Vol. 69, No. 5 September 1962

Psychological Review

RICHARD L. SOLOMON, Editor
University of Pennsylvania

CONTENTS

ARTICLES

Cognitive, Social, and Physiological Determinants of Emotional State
STANLEY SCHACHTER AND JEROME E. SINGER 379

On the Quantification of Miller's Conflict Theory
NORMAN H. ANDERSON 400

From Association to Structure.....GEORGE MANDLER 415

Stimulus-Intensity Effects in Response Evocation..R. A. CHAMPION 428

Conditioning through Vicarious Instigation...SEYMOUR M. BERGER 450

Conditional-Outcome Choice Behavior in Rats...FRANK A. LOGAN 467

PUBLISHED BIMONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

Editorial Consultants

ROBERT P. ABELSON
ALBERT BANDURA
GORDON H. BOWER
RICHARD CHRISTIE
RUSSELL M. CHURCH
JAMES DEESE
CARL P. DUNCAN
J. A. EASTERBROOK
JOHN GAITO
DAVID A. GRANT
ECKHARD H. HESS
R. J. HERRNSTEIN
WINFRED F. HILL
JULIAN HOCHBERG
J. MCV. HUNT
HOWARD H. KENDLER
TRACY S. KENDLER
WILLIAM KESSEN
IRVING MALTZMAN
SAMUEL J. MESSICK
CONRAD G. MUELLER
HELEN PEAK
CHARLES C. PERKINS, JR.
LEO J. POSTMAN
MALCOLM G. PRESTON
K. EDWARD RENNER
MARK R. ROSENZWEIG
HAROLD SCHLOSBERG
KENNETH W. SPENCE
ROGER W. SPERRY
ARTHUR W. STAATS
SAUL H. STERNBERG
RENATO TAGIURI
HANS-LUKAS TEUBER
BENTON J. UNDERWOOD

(A large group of psychologists has kindly agreed to assist in the editing of manuscripts submitted to the *Psychological Review*. In each issue the Editor will list in the *Review* those editorial consultants who have helped him process manuscripts for that issue, so that their efforts may be gratefully acknowledged.)

The *Psychological Review* is devoted to articles of theoretical significance to any area of scientific endeavor in psychology. Except for occasional articles solicited by the Editor, manuscripts exceeding 7,500 words will not be accepted. Ordinarily manuscripts which consist primarily of original reports of research should be submitted to other journals. However, reports of original research may be included in a major theoretical article which attempts to integrate several related original studies.

Because of the large number of manuscripts submitted, there is an inevitable publication lag of several months. Authors may avoid this delay if they are prepared to pay the costs of publishing their own articles; the appearance of articles by other contributors is not thereby delayed.

Manuscripts must be accompanied by an abstract of 100 to 120 words typed on a separate sheet of paper. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstract may be obtained from the Editor or from the APA Central Office.

All manuscripts must be submitted in duplicate.

Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors should retain a copy of the manuscript, in case there is loss in the mail. Manuscripts should be addressed to the Editor, Richard L. Solomon, Psychology Department, College Hall, University of Pennsylvania, Philadelphia 4, Pennsylvania.

HELEN ORR, Managing Editor
ELIZABETH S. REED, Advertising Manager
SHARON STONE, Editorial Assistant

PUBLISHED BIMONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Prince and Lemon Streets, Lancaster, Pa.
1333 Sixteenth St. N.W., Washington 6, D. C.

$10.00 volume $2.00 issue

Second-class postage paid at Lancaster, Pa.

Send all communications, including address changes, to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must arrive by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1962 by the American Psychological Association, Inc.

VOL. 69, NO. 5 SEPTEMBER 1962

PSYCHOLOGICAL REVIEW

COGNITIVE, SOCIAL, AND PHYSIOLOGICAL DETERMINANTS OF EMOTIONAL STATE[1]

STANLEY SCHACHTER AND JEROME E. SINGER
Columbia University *Pennsylvania State University*

The problem of which cues, internal or external, permit a person to label and identify his own emotional state has been with us since the days that James (1890) first tendered his doctrine that "the bodily changes follow directly the perception of the exciting fact, and that our feeling of the same changes as they occur *is* the emotion" (p. 449). Since we are aware of a variety of feeling and emotion states, it should follow from James' proposition that the various emotions will be accompanied by a variety of differentiable bodily states. Following James' pronouncement, a formidable number of studies were undertaken in search of the physiological differentiators of the emotions. The results, in these early days, were almost uniformly negative. All of the emotional states experimentally manipulated were characterized by a general pattern of excitation of the sympathetic nervous system but there appeared to be no clear-cut physiological discriminators of the various emotions. This pattern of results was so consistent from experiment to experiment that Cannon (1929) offered, as one of the crucial criticisms of the James-Lange theory, the fact that "the same visceral changes occur in very different emotional states and in non-emotional states" (p. 351).

More recent work, however, has given some indication that there may be differentiators. Ax (1953) and Schachter (1957) studied fear and anger. On a large number of indices both of these states were characterized by a similarly high level of autonomic activation but on several indices they did differ in the degree of activation. Wolf and Wolff (1947) studied a subject with a gastric fistula and were able to distinguish two patterns in the physiological responses of the stomach wall. It should be noted, though, that for many months they studied their subject during and following a great variety of moods and emotions and were able to distinguish only two patterns.

Whether or not there are physiological distinctions among the various emotional states must be considered an open

[1] This experiment is part of a program of research on cognitive and physiological determinants of emotional state which is being conducted at the Department of Social Psychology at Columbia University under PHS Research Grant M-2584 from the National Institute of Mental Health, United States Public Health Service. This experiment was conducted at the Laboratory for Research in Social Relations at the University of Minnesota.

The authors wish to thank Jean Carlin and Ruth Hase, the physicians in the study, and Bibb Latané and Leonard Weller who were the paid participants.

question. Recent work might be taken to indicate that such differences are at best rather subtle and that the variety of emotion, mood, and feeling states are by no means matched by an equal variety of visceral patterns.

This rather ambiguous situation has led Ruckmick (1936), Hunt, Cole, and Reis (1958), Schachter (1959) and others to suggest that cognitive factors may be major determinants of emotional states. Granted a general pattern of sympathetic excitation as characteristic of emotional states, granted that there may be some differences in pattern from state to state, it is suggested that one labels, interprets, and identifies this stirred-up state in terms of the characteristics of the precipitating situation and one's apperceptive mass. This suggests, then, that an emotional state may be considered a function of a state of physiological arousal [2] and of a cognition appropriate to this state of arousal. The cognition, in a sense, exerts a steering function. Cognitions arising from the immediate situation as interpreted by past experience provide the framework within which one understands and labels his feelings. It is the cognition which determines whether the state of physiological arousal will be labeled as "anger," "joy," "fear," or whatever.

In order to examine the implications of this formulation let us consider the fashion in which these two elements, a state of physiological arousal and cognitive factors, would interact in a variety of situations. In most emotion inducing situations, of course, the two factors are completely interrelated. Imagine a man walking alone down a dark alley, a figure with a gun suddenly appears. The perception-cognition "figure with a gun" in some fashion initiates a state of physiological arousal; this state of arousal is interpreted in terms of knowledge about dark alleys and guns and the state of arousal is labeled "fear." Similarly a student who unexpectedly learns that he has made Phi Beta Kappa may experience a state of arousal which he will label "joy."

Let us now consider circumstances in which these two elements, the physiological and the cognitive, are, to some extent, independent. First, is the state of physiological arousal alone sufficient to induce an emotion? Best evidence indicates that it is not. Marañon [3] (1924), in a fascinating study, (which was replicated by Cantril & Hunt, 1932, and Landis & Hunt, 1932) injected 210 of his patients with the sympathomimetic agent adrenalin and then simply asked them to introspect. Seventy-one percent of his subjects simply reported their physical symptoms with no emotional overtones; 29% of the subjects responded in an apparently emotional fashion. Of these the great majority described their feelings in a fashion that Marañon labeled "cold" or "as if" emotions, that is, they made statements such as "I feel *as if* I were afraid" or "*as if* I were awaiting a great happiness." This is a sort of emotional "déjà vu" experience; these subjects are neither happy nor afraid, they feel "as if" they were. Finally a very few cases apparently reported a genuine emotional experience. However, in order to produce this reaction in most of these few cases, Marañon (1924) points out:

[2] Though our experiments are concerned exclusively with the physiological changes produced by the injection of adrenalin, which appear to be primarily the result of sympathetic excitation, the term physiological arousal is used in preference to the more specific "excitation of the sympathetic nervous system" because there are indications, to be discussed later, that this formulation is applicable to a variety of bodily states.

[3] Translated copies of Marañon's (1924) paper may be obtained by writing to the senior author.

> One must suggest a memory with strong affective force but not so strong as to produce an emotion in the normal state. For example, in several cases we spoke to our patients before the injection of their sick children or dead parents and they responded calmly to this topic. The same topic presented later, during the adrenal commotion, was sufficient to trigger emotion. This adrenal commotion places the subject in a situation of 'affective imminence' (pp. 307–308).

Apparently, then, to produce a genuinely emotional reaction to adrenalin, Marañon was forced to provide such subjects with an appropriate cognition.

Though Marañon (1924) is not explicit on his procedure, it is clear that his subjects knew that they were receiving an injection and in all likelihood knew that they were receiving adrenalin and probably had some order of familiarity with its effects. In short, though they underwent the pattern of sympathetic discharge common to strong emotional states, at the same time they had a completely appropriate cognition or explanation as to why they felt this way. This, we would suggest, is the reason so few of Marañon's subjects reported any emotional experience.

Consider now a person in a state of physiological arousal for which no immediately explanatory or appropriate cognitions are available. Such a state could result were one covertly to inject a subject with adrenalin or, unknown to him, feed the subject a sympathomimetic drug such as ephedrine. Under such conditions a subject would be aware of palpitations, tremor, face flushing, and most of the battery of symptoms associated with a discharge of the sympathetic nervous system. In contrast to Marañon's (1924) subjects he would, at the same time, be utterly unaware of why he felt this way. What would be the consequence of such a state?

Schachter (1959) has suggested that precisely such a state would lead to the arousal of "evaluative needs" (Festinger, 1954), that is, pressures would act on an individual in such a state to understand and label his bodily feelings. His bodily state grossly resembles the condition in which it has been at times of emotional excitement. How would he label his present feelings? It is suggested, of course, that he will label his feelings in terms of his knowledge of the immediate situation.[4] Should he at the time be with a beautiful woman he might decide that he was wildly in love or sexually excited. Should he be at a gay party, he might, by comparing himself to others, decide that he was extremely happy and euphoric. Should he be arguing with his wife, he might explode in fury and hatred. Or, should the situation be completely inappropriate he could decide that he was excited about something that had recently happened to him or, simply, that he was sick. In any case, it is our basic assumption that emotional states are a function of the interaction of such cognitive factors with a state of physiological arousal.

This line of thought, then, leads to the following propositions:

1. Given a state of physiological arousal for which an individual has no immediate explanation, he will "label" this state and describe his feelings in terms of the cognitions available to him. To the extent that cognitive factors are potent determiners of emotional states, it could be anticipated that precisely the same state of physiological arousal could be labeled "joy" or "fury" or "jealousy" or any of a great diversity

[4] This suggestion is not new for several psychologists have suggested that situational factors should be considered the chief differentiators of the emotions. Hunt, Cole, and Reis (1958) probably make this point most explicitly in their study distinguishing among fear, anger, and sorrow in terms of situational characteristics.

of emotional labels depending on the cognitive aspects of the situation.

2. Given a state of physiological arousal for which an individual has a completely appropriate explanation (e.g., "I feel this way because I have just received an injection of adrenalin") no evaluative needs will arise and the individual is unlikely to label his feelings in terms of the alternative cognitions available.

Finally, consider a condition in which emotion inducing cognitions are present but there is no state of physiological arousal. For example, an individual might be completely aware that he is in great danger but for some reason (drug or surgical) remain in a state of physiological quiescence. Does he experience the emotion "fear"? Our formulation of emotion as a joint function of a state of physiological arousal and an appropriate cognition, would, of course, suggest that he does not, which leads to our final proposition.

3. Given the same cognitive circumstances, the individual will react emotionally or describe his feelings as emotions only to the extent that he experiences a state of physiological arousal.[5]

PROCEDURE

The experimental test of these propositions requires (*a*) the experimental manipulation of a state of physiological arousal, (*b*) the manipulation of the extent to which the subject has an appropriate or proper explanation of his bodily state, and (*c*) the creation of situations from which explanatory cognitions may be derived.

In order to satisfy the first two experimental requirements, the experiment was cast in the framework of a study of the effects of vitamin supplements on vision. As soon as a subject arrived, he was taken to a private room and told by the experimenter:

> In this experiment we would like to make various tests of your vision. We are particularly interested in how certain vitamin compounds and vitamin supplements affect the visual skills. In particular, we want to find out how the vitamin compound called 'Suproxin' affects your vision.
>
> What we would like to do, then, if we can get your permission, is to give you a small injection of Suproxin. The injection itself is mild and harmless; however, since some people do object to being injected we don't want to talk you into anything. Would you mind receiving a Suproxin injection?

If the subject agrees to the injection (and all but 1 of 185 subjects did) the experimenter continues with instructions we shall describe shortly, then leaves the room. In a few minutes a physician enters the room, briefly repeats the experimenter's instructions, takes the subject's pulse and then injects him with Suproxin.

Depending upon condition, the subject receives one of two forms of Suproxin—epinephrine or a placebo.

Epinephrine or adrenalin is a sympathomimetic drug whose effects, with minor exceptions, are almost a perfect mimicry of a discharge of the sympathetic nervous system. Shortly after injection systolic blood pressure increases markedly, heart rate increases somewhat, cutaneous blood flow decreases, while muscle and cerebral blood flow increase, blood sugar and lactic acid concentration increase, and respiration rate increases slightly. As far as the subject is concerned the major subjective symptoms are palpitation, tremor, and sometimes a feeling of flushing and accelerated breathing. With a subcutaneous injection (in the dosage administered to our subjects), such effects usually begin within 3–5 minutes of injection and last anywhere from 10 minutes to an hour. For most subjects these effects are dissipated within 15–20 minutes after injection.

Subjects receiving epinephrine received a subcutaneous injection of ½ cubic centimeter of a 1 : 1000 solution of Winthrop Laboratory's Suprarenin, a saline solution of epinephrine bitartrate.

Subjects in the placebo condition received a subcutaneous injection of ½ cubic centimeter of saline solution. This is, of course, com-

[5] In his critique of the James-Lange theory of emotion, Cannon (1929) also makes the point that sympathectomized animals and patients do seem to manifest emotional behavior. This criticism is, of course, as applicable to the above proposition as it was to the James-Lange formulation. We shall discuss the issues involved in later papers.

pletely neutral material with no side effects at all.

Manipulating an Appropriate Explanation

By "appropriate" we refer to the extent to which the subject has an authoritative, unequivocal explanation of his bodily condition. Thus, a subject who had been informed by the physician that as a direct consequence of the injection he would feel palpitations, tremor, etc. would be considered to have a completely appropriate explanation. A subject who had been informed only that the injection would have no side effects would have no appropriate explanation of his state. This dimension of appropriateness was manipulated in three experimental conditions which shall be called: Epinephrine Informed (Epi Inf), Epinephrine Ignorant (Epi Ign), and Epinephrine Misinformed (Epi Mis).

Immediately after the subject had agreed to the injection and before the physician entered the room, the experimenter's spiel in each of these conditions went as follows:

> *Epinephrine Informed.* I should also tell you that some of our subjects have experienced side effects from the Suproxin. These side effects are transitory, that is, they will only last for about 15 or 20 minutes. What will probably happen is that your hand will start to shake, your heart will start to pound, and your face may get warm and flushed. Again these are side effects lasting about 15 or 20 minutes.

While the physician was giving the injection, she told the subject that the injection was mild and harmless and repeated this description of the symptoms that the subject could expect as a consequence of the shot. In this condition, then, subjects have a completely appropriate explanation of their bodily state. They know precisely what they will feel and why.

> *Epinephrine Ignorant.* In this condition, when the subject agreed to the injection, the experimenter said nothing more relevant to side effects and simply left the room. While the physician was giving the injection, she told the subject that the injection was mild and harmless and would have no side effects. In this condition, then, the subject has no experimentally provided explanation for his bodily state.

> *Epinephrine Misinformed.* I should also tell you that some of our subjects have experienced side effects from the Suproxin. These side effects are transitory, that is, they will only last for about 15 or 20 minutes. What will probably happen is that your feet will feel numb, you will have an itching sensation over parts of your body, and you may get a slight headache. Again these are side effects lasting 15 or 20 minutes.

And again, the physician repeated these symptoms while injecting the subject.

None of these symptoms, of course, are consequences of an injection of epinephrine and, in effect, these instructions provide the subject with a completely inappropriate explanation of his bodily feelings. This condition was introduced as a control condition of sorts. It seemed possible that the description of side effects in the Epi Inf condition might turn the subject introspective, self-examining, possibly slightly troubled. Differences on the dependent variable between the Epi Inf and Epi Ign conditions might, then, be due to such factors rather than to differences in appropriateness. The false symptoms in the Epi Mis condition should similarly turn the subject introspective, etc., but the instructions in this condition do not provide an appropriate explanation of the subject's state.

Subjects in all of the above conditions were injected with epinephrine. Finally, there was a placebo condition in which subjects, who were injected with saline solution, were given precisely the same treatment as subjects in the Epi Ign condition.

Producing an Emotion Inducing Cognition

Our initial hypothesis has suggested that given a state of physiological arousal for which the individual has no adequate explanation, cognitive factors can lead the individual to describe his feelings with any of a diversity of emotional labels. In order to test this hypothesis, it was decided to manipulate emotional states which can be considered quite different—euphoria and anger.

There are, of course, many ways to induce such states. In our own program of research, we have concentrated on social determinants of emotional states and have been able to demonstrate in other studies that people do evaluate their own feelings by comparing themselves with others around them (Schachter 1959; Wrightsman 1960). In this experiment we have attempted again to manipulate emotional state by social means. In one set of conditions, the subject is placed together with a stooge who has been trained

to act euphorically. In a second set of conditions the subject is with a stooge trained to act in an angry fashion.

Euphoria

Immediately [6] after the subject had been injected, the physician left the room and the experimenter returned with a stooge whom he introduced as another subject, then said:

> Both of you have had the Suproxin shot and you'll both be taking the same tests of vision. What I ask you to do now is just wait for 20 minutes. The reason for this is simply that we have to allow 20 minutes for the Suproxin to get from the injection site into the bloodstream. At the end of 20 minutes when we are certain that most of the Suproxin has been absorbed into the bloodstream, we'll begin the tests of vision.

The room in which this was said had been deliberately put into a state of mild disarray. As he was leaving, the experimenter apologetically added:

> The only other thing I should do is to apologize for the condition of the room. I just didn't have time to clean it up. So, if you need any scratch paper or rubber bands or pencils, help yourself. I'll be back in 20 minutes to begin the vision tests.

As soon as the experimenter had left, the stooge introduced himself again, made a series of standard icebreaker comments, and then launched his routine. For observation purposes, the stooge's act was broken into a series of standard units, demarcated by a change in activity or a standard comment. In sequence, the units of the stooge's routine were the following:

1. Stooge reaches for a piece of paper and starts doodling saying, "They said we could use this for scratch, didn't they?" He doodles a fish for some 30 seconds, then says:

2. "This scrap paper isn't even much good for doodling" and crumples paper and attempts to throw it into wastebasket in far corner of the room. He misses but this leads him into a "basketball game." He crumples up other sheets of paper, shoots a few baskets, says "Two points" occasionally. He gets up and does a jump shot saying, "The old jump shot is really on today."

3. If the subject has not joined in, the stooge throws a paper basketball to the subject saying, "Here, you try it."

4. Stooge continues his game saying, "The trouble with paper basketballs is that you don't really have any control."

5. Stooge continues basketball, then gives it up saying, "This is one of my good days. I feel like a kid again. I think I'll make a plane." He makes a paper airplane saying, "I guess I'll make one of the longer ones."

6. Stooge flies plane. Gets up and retrieves plane. Flies again, etc.

7. Stooge throws plane at subject.

8. Stooge, flying plane, says, "Even when I was a kid, I was never much good at this."

9. Stooge tears off part of plane saying, "Maybe this plane can't fly but at least it's good for something." He wads up paper and making a slingshot of a rubber band begins to shoot the paper.

10. Shooting, the stooge says, "They [paper ammunition] really go better if you make them long. They don't work right if you wad them up."

11. While shooting, stooge notices a sloppy pile of manila folders on a table. He builds a tower of these folders, then goes to the opposite end of the room to shoot at the tower.

12. He misses several times, then hits and cheers as the tower falls. He goes over to pick up the folders.

13. While picking up, he notices, behind a portable blackboard, a pair of hula hoops which have been covered with black tape with a few wires sticking out of the tape. He reaches for these, taking one for himself and putting the other aside but within reaching distance of the subject. The stooge tries the hula hoop, saying, "This isn't as easy as it looks."

14. Stooge twirls hoop wildly on arm, saying, "Hey, look at this—this is great."

15. Stooge replaces the hula hoop and sits down with his feet on the table. Shortly thereafter the experimenter returns to the room.

[6] It was, of course, imperative that the sequence with the stooge begin before the subject felt his first symptoms for otherwise the subject would be virtually forced to interpret his feelings in terms of events preceding the stooge's entrance. Pretests had indicated that, for most subjects, epinphrine-caused symptoms began within 3–5 minutes after injection. A deliberate attempt was made then to bring in the stooge within 1 minute after the subject's injection.

This routine was completely standard, though its pace, of course, varied depending upon the subject's reaction, the extent to which he entered into this bedlam and the extent to which he initiated activities of his own. The only variations from this standard routine were those forced by the subject. Should the subject originate some nonsense of his own and request the stooge to join in, he would do so. And, he would, of course, respond to any comments initiated by the subject.

Subjects in each of the three "appropriateness" conditions and in the placebo condition were submitted to this setup. The stooge, of course, never knew in which condition any particular subject fell.

Anger

Immediately after the injection, the experimenter brought a stooge into the subject's room, introduced the two and after explaining the necessity for a 20 minute delay for "the Suproxin to get from the injection site into the bloodstream" he continued, "We would like you to use these 20 minutes to answer these questionnaires." Then handing out the questionnaires, he concludes with, "I'll be back in 20 minutes to pick up the questionnaires and begin the tests of vision."

Before looking at the questionnaire, the stooge says to the subject,

> I really wanted to come for an experiment today, but I think it's unfair for them to give you shots. At least, they should have told us about the shots when they called us; you hate to refuse, once you're here already.

The questionnaires, five pages long, start off innocently requesting face sheet information and then grow increasingly personal and insulting. The stooge, sitting directly opposite the subject, paces his own answers so that at all times subject and stooge are working on the same question. At regular points in the questionnaire, the stooge makes a series of standardized comments about the questions. His comments start off innocently enough, grow increasingly querulous, and finally he ends up in a rage. In sequence, he makes the following comments.

1. Before answering any items, he leafs quickly through the questionnaire saying, "Boy, this is a long one."

2. Question 7 on the questionnaire requests, "List the foods that you would eat in a typical day." The stooge comments, "Oh for Pete's sake, what did I have for breakfast this morning?"

3. Question 9 asks, "Do you ever hear bells? ———. How often? ———." The stooge remarks, "Look at Question 9. How ridiculous can you get? I hear bells every time I change classes."

4. Question 13 requests, "List the childhood diseases you have had and the age at which you had them" to which the stooge remarks, "I get annoyed at this childhood disease question. I can't remember what childhood diseases I had, and especially at what age. Can you?"

5. Question 17 asks "What is your father's average annual income?" and the stooge says, "This really irritates me. It's none of their business what my father makes. I'm leaving that blank."

6. Question 25 presents a long series of items such as "Does not bathe or wash regularly," "Seems to need psychiatric care," etc. and requests the respondent to write down for which member of his immediate family each item seems most applicable. The question specifically prohibits the answer "None" and each item must be answered. The stooge says, "I'll be damned if I'll fill out Number 25. 'Does not bathe or wash regularly'—that's a real insult." He then angrily crosses out the entire item.

7. Question 28 reads:

"How many times each week do you have sexual intercourse?" 0–1 —— 2–3 —— 4–6 —— 7 and over ——. The stooge bites out, "The hell with it! I don't have to tell them all this."

8. The stooge sits sullenly for a few moments then he rips up his questionnaire, crumples the pieces and hurls them to the floor, saying, "I'm not wasting any more time. I'm getting my books and leaving" and he stamps out of the room.

9. The questionnaire continues for eight more questions ending with: "With how many men (other than your father) has your mother had extramarital relationships?"

4 and under ——: 5–9 ——: 10 and over ——.

Subjects in the Epi Ign, Epi Inf and Placebo conditions were run through this "anger" inducing sequence. The stooge, again, did not know to which condition the subject had been assigned.

In summary, this is a seven condition experiment which, for two different emotional states, allows us (*a*) to evaluate the effects of "appropriateness" on emotional inducibility

and (*b*) to begin to evaluate the effects of sympathetic activation on emotional inducibility. In schematic form the conditions are the following:

EUPHORIA	ANGER
Epi Inf	Epi Inf
Epi Ign	Epi Ign
Epi Mis	Placebo
Placebo	

The Epi Mis condition was not run in the Anger sequence. This was originally conceived as a control condition and it was felt that its inclusion in the Euphoria conditions alone would suffice as a means of evaluating the possible artifactual effect of the Epi Inf instructions.

Measurement

Two types of measures of emotional state were obtained. Standardized observation through a one-way mirror was the technique used to assess the subject's behavior. To what extent did he act euphoric or angry? Such behavior can be considered in a way as a "semiprivate" index of mood for as far as the subject was concerned, his emotional behavior could be known only to the other person in the room—presumably another student. The second type of measure was self-report in which, on a variety of scales, the subject indicated his mood of the moment. Such measures can be considered "public" indices of mood for they would, of course, be available to the experimenter and his associates.

Observation

Euphoria. For each of the first 14 units of the stooge's standardized routine an observer kept a running chronicle of what the subject did and said. For each unit the observer coded the subject's behavior in one or more of the following categories:

Category 1: Joins in activity. If the subject entered into the stooge's activities, e.g., if he made or flew airplanes, threw paper basketballs, hula hooped, etc., his behavior was coded in this category.

Category 2: Initiates new activity. A subject was so coded if he gave indications of creative euphoria, that is, if, on his own, he initiated behavior outside of the stooge's routine. Instances of such behavior would be the subject who threw open the window and, laughing, hurled paper basketballs at passersby; or, the subject who jumped on a table and spun one hula hoop on his leg and the other on his neck.

Categories 3 and 4: Ignores or watches stooge. Subjects who paid flatly no attention to the stooge or who, with or without comment, simply watched the stooge without joining in his activity were coded in these categories.

For any particular unit of behavior, the subject's behavior was coded in one or more of these categories. To test reliability of coding two observers independently coded two experimental sessions. The observers agreed completely on the coding of 88% of the units.

Anger. For each of the units of stooge behavior, an observer recorded the subject's responses and coded them according to the following category scheme:

Category 1: Agrees. In response to the stooge the subject makes a comment indicating that he agrees with the stooge's standardized comment or that he, too, is irked by a particular item on the questionnaire. For example, a subject who responded to the stooge's comment on the "father's income" question by saying, "I don't like that kind of personal question either" would be so coded (scored +2).

Category 2: Disagrees. In response to the stooge's comment, the subject makes a comment which indicates that he disagrees with the stooge's meaning or mood; e.g., in response to the stooge's comment on the "father's income" question, such a subject might say, "Take it easy, they probably have a good reason for wanting the information" (scored −2).

Category 3: Neutral. A noncommittal or irrelevant response to the stooge's remark (scored 0).

Category 4: Initiates agreement or disagreement. With no instigation by the stooge, a subject, so coded, would have volunteered a remark indicating that he felt the same way or, alternatively, quite differently than the stooge. Examples would be "Boy I hate this kind of thing" or "I'm enjoying this" (scored +2 or −2).

Category 5: Watches. The subject makes no verbal response to the stooge's comment but simply looks directly at him (scored 0).

Category 6: Ignores. The subject makes no verbal response to the stooge's comment nor does he look at him; the subject, paying no attention at all to the stooge, simply works at his own questionnaire (scored −1).

A subject was scored in one or more of these categories for each unit of stooge behavior. To test reliability, two observers independently coded three experimental sessions. In order to get a behavioral index of anger, observation protocol was scored according to the values presented in parentheses

after each of the above definitions of categories. In a unit-by-unit comparison, the two observers agreed completely on the scoring of 71% of the units jointly observed. The scores of the two observers differed by a value of 1 or less for 88% of the units coded and in not a single case did the two observers differ in the direction of their scoring of a unit.

Self Report of Mood and Physical Condition

When the subject's session with the stooge was completed, the experimenter returned to the room, took pulses and said:

> Before we proceed with the vision tests, there is one other kind of information which we must have. We have found, as you can probably imagine, that there are many things beside Suproxin that affect how well you see in our tests. How hungry you are, how tired you are, and even the mood you're in at the time—whether you feel happy or irritated at the time of testing will affect how well you see. To understand the data we collect on you, then, we must be able to figure out which effects are due to causes such as these and which are caused by Suproxin.
>
> The only way we can get such information about your physical and emotional state is to have you tell us. I'll hand out these questionnaires and ask you to answer them as accurately as possible. Obviously, our data on the vision tests will only be as accurate as your description of your mental and physical state.

In keeping with this spiel, the questionnaire that the experimenter passed out contained a number of mock questions about hunger, fatigue, etc., as well as questions of more immediate relevance to the experiment. To measure mood or emotional state the following two were the crucial questions:

1. How irritated, angry or annoyed would you say you feel at present?

I don't feel at all irritated or angry (0)	I feel a little irritated and angry (1)	I feel quite irritated and angry (2)	I feel very irritated and angry (3)	I feel extremely irritated and angry (4)

2. How good or happy would you say you feel at present?

I don't feel at all happy or good (0)	I feel a little happy and good (1)	I feel quite happy and good (2)	I feel very happy and good (3)	I feel extremely happy and good (4)

To measure the physical effects of epinephrine and determine whether or not the injection had been successful in producing the necessary bodily state, the following questions were asked:

1. Have you experienced any palpitation (consciousness of your own heart beat)?

Not at all (0)	A slight amount (1)	A moderate amount (2)	An intense amount (3)

2. Did you feel any tremor (involuntary shaking of the hands, arms or legs)?

Not at all (0)	A slight amount (1)	A moderate amount (2)	An intense amount (3)

To measure possible effects of the instructions in the Epi Mis condition, the following questions were asked:

1. Did you feel any numbness in your feet?
2. Did you feel any itching sensation?
3. Did you experience any feeling of headache?

To all three of these questions was attached a four-point scale running from "Not at all" to "An intense amount."

In addition to these scales, the subjects were asked to answer two open-end questions on other physical or emotional sensations they may have experienced during the experimental session. A final measure of bodily state was pulse rate which was taken by the physician or the experimenter at two times—immediately before the injection and immediately after the session with the stooge.

When the subjects had completed these questionnaires, the experimenter announced that the experiment was over, explained the deception and its necessity in detail, answered any questions, and swore the subjects to secrecy. Finally, the subjects answered a brief questionnaire about their experiences, if any, with adrenalin and their previous knowledge or suspicion of the experimental setup. There was no indication that any of the subjects had known about the experiment beforehand but 11 subjects were so extremely suspicious of some crucial feature of the experiment that their data were automatically discarded.

Subjects

The subjects were all male, college students taking classes in introductory psychology at the University of Minnesota. Some 90% of the students in these classes volunteer for a subject pool for which they receive two extra points on their final exam for every

hour that they serve as experimental subjects. For this study the records of all potential subjects were cleared with the Student Health Service in order to insure that no harmful effects would result from the injections.

Evaluation of the Experimental Design

The ideal test of our propositions would require circumstances which our experiment is far from realizing. First, the proposition that: "A state of physiological arousal for which an individual has no immediate explanation will lead him to label this state in terms of the cognitions available to him" obviously requires conditions under which the subject does not and cannot have a proper explanation of his bodily state. Though we toyed with such fantasies as ventilating the experimental room with vaporized adrenalin, reality forced us to rely on the disguised injection of Suproxin—a technique which was far from ideal for no matter what the experimenter told them, some subjects would inevitably attribute their feelings to the injection. To the extent that subjects did so, differences between the several appropriateness conditions should be attenuated.

Second, the proposition that: "Given the same cognitive circumstances the individual will react emotionally only to the extent that he experiences a state of physiological arousal" requires for its ideal test the manipulation of states of physiological arousal and of physiological quiescence. Though there is no question that epinephrine effectively produces a state of arousal, there is also no question that a placebo does not prevent physiological arousal. To the extent that the experimental situation effectively produces sympathetic stimulation in placebo subjects, the proposition is difficult to test, for such a factor would attenuate differences between epinephrine and placebo subjects.

Both of these factors, then, can be expected to interfere with the test of our several propositions. In presenting the results of this study, we shall first present condition by condition results and then evaluate the effect of these two factors on experimental differences.

RESULTS

Effects of the Injections on Bodily State

Let us examine first the success of the injections at producing the bodily state required to examine the propositions at test. Does the injection of epinephrine produce symptoms of sympathetic discharge as compared with the placebo injection? Relevant data are presented in Table 1 where it can be immediately seen that on all items subjects who were in epinephrine conditions show considerably more evidence of sympathetic activation than do subjects in placebo conditions. In all epinephrine conditions pulse rate increases significantly when compared with the decrease characteristic of the placebo conditions. On the scales it is clear that epinephrine subjects experi-

TABLE 1

THE EFFECTS OF THE INJECTIONS ON BODILY STATE

Condition	*N*	Pulse		Self-rating of				
		Pre	Post	Palpitation	Tremor	Numbness	Itching	Headache
Euphoria								
Epi Inf	27	85.7	88.6	1.20	1.43	0	0.16	0.32
Epi Ign	26	84.6	85.6	1.83	1.76	0.15	0	0.55
Epi Mis	26	82.9	86.0	1.27	2.00	0.06	0.08	0.23
Placebo	26	80.4	77.1	0.29	0.21	0.09	0	0.27
Anger								
Epi Inf	23	85.9	92.4	1.26	1.41	0.17	0	0.11
Epi Ign	23	85.0	96.8	1.44	1.78	0	0.06	0.21
Placebo	23	84.5	79.6	0.59	0.24	0.14	0.06	0.06

ence considerably more palpitation and tremor than do placebo subjects. In all possible comparisons on these symptoms, the mean scores of subjects in any of the epinephrine conditions are greater than the corresponding scores in the placebo conditions at better than the .001 level of significance. Examination of the absolute values of these scores makes it quite clear that subjects in epinephrine conditions were, indeed, in a state of physiological arousal, while most subjects in placebo conditions were in a relative state of physiological quiescence.

The epinephrine injection, of course, did not work with equal effectiveness for all subjects; indeed for a few subjects it did not work at all. Such subjects reported almost no palpitation or tremor, showed no increase in pulse and described no other relevant physical symptoms. Since for such subjects the necessary experimental conditions were not established, they were automatically excluded from the data and all further tabular presentations will not include such subjects. Table 1, however, does include the data of these subjects. There were four such subjects in euphoria conditions and one of them in anger conditions.

In order to evaluate further data on Epi Mis subjects it is necessary to note the results of the "numbness," "itching," and "headache" scales also presented in Table 1. Clearly the subjects in the Epi Mis condition do not differ on these scales from subjects in any of the other experimental conditions.

Effects of the Manipulations on Emotional State

Euphoria: Self-report. The effects of the several manipulations on emotional state in the euphoria conditions are presented in Table 2. The scores recorded in this table are derived, for each subject, by subtracting the value of the point he checks on the irritation scale from the value of the point he checks on the happiness scale. Thus, if a subject were to check the point "I feel a little irritated and angry" on the irritation scale and the point "I feel very happy and good" on the happiness scale, his score would be +2. The higher the positive value, the happier and better the subject reports himself as feeling. Though we employ an index for expositional simplicity, it should be noted that the two components of the index each yield results completely consistent with those obtained by use of this index.

Let us examine first the effects of the appropriateness instructions. Comparison of the scores for the Epi Mis and Epi Inf conditions makes it immediately clear that the experimental differences are not due to artifacts resulting from the informed instructions. In both conditions the subject was warned to expect a variety of symptoms as a consequence of the injection. In the Epi Mis condition, where the symptoms were inappropriate to the subject's bodily state the self-report score is almost twice that in the Epi Inf condition where the symptoms were completely appropriate to the subject's bodily state. It is reasonable, then, to attribute differences between informed subjects and those in other conditions to differences in manipulated appropriateness rather than to artifacts such as introspectiveness or self-examination.

It is clear that, consistent with expectations, subjects were more susceptible to the stooge's mood and consequently more euphoric when they had no explanation of their own bodily states than when they did. The means of both the Epi Ign and Epi Mis conditions are considerably greater than the mean of the Epi Inf condition.

It is of interest to note that Epi Mis subjects are somewhat more euphoric

TABLE 2

SELF-REPORT OF EMOTIONAL STATE IN THE EUPHORIA CONDITIONS

Condition	N	Self-Report scales	Comparison	p*
Epi Inf	25	0.98	Epi Inf vs. Epi Mis	<.01
Epi Ign	25	1.78	Epi Inf vs. Epi Ign	.02
Epi Mis	25	1.90	Placebo vs. Epi Mis, Ign, or Inf	ns
Placebo	26	1.61		

All p values reported throughout paper are two-tailed.

than are Epi Ign subjects. This pattern repeats itself in other data shortly to be presented. We would attribute this difference to differences in the appropriateness dimension. Though, as in the Epi Ign condition, a subject is not provided with an explanation of his bodily state, it is, of course, possible that he will provide one for himself which is not derived from his interaction with the stooge. Most reasonably he could decide for himself that he feels this way because of the injection. To the extent that he does so he should be less susceptible to the stooge. It seems probable that he would be less likely to hit on such an explanation in the Epi Mis condition than in the Epi Ign condition for in the Epi Mis condition both the experimenter and the doctor have told him that the effects of the injection would be quite different from what he actually feels. The effect of such instructions is probably to make it more difficult for the subject himself to hit on the alternative explanation described above. There is some evidence to support this analysis. In open-end questions in which subjects described their own mood and state, 28% of the subjects in the Epi Ign condition made some connection between the injection and their bodily state compared with the 16% of subjects in the Epi Mis condition who did so. It could be considered, then, that these three conditions fall along a dimension of appropriateness, with the Epi Inf condition at one extreme and the Epi Mis condition at the other.

Comparing the placebo to the epinephrine conditions, we note a pattern which will repeat itself throughout the data. Placebo subjects are less euphoric than either Epi Mis or Epi Ign subjects but somewhat more euphoric than Epi Inf subjects. These differences are not, however, statistically significant. We shall consider the epinephrine-placebo comparisons in detail in a later section of this paper following the presentation of additional relevant data. For the moment, it is clear that, by self-report manipulating appropriateness has had a very strong effect on euphoria.

Behavior. Let us next examine the extent to which the subject's behavior was affected by the experimental manipulations. To the extent that his mood has been affected, one should expect that the subject will join in the stooge's whirl of manic activity and initiate similar activities of his own. The relevant data are presented in Table 3. The column labeled "Activity

TABLE 3

BEHAVIORAL INDICATIONS OF EMOTIONAL STATE IN THE EUPHORIA CONDITIONS

Condition	N	Activity index	Mean number of acts initiated
Epi Inf	25	12.72	.20
Epi Ign	25	18.28	.56
Epi Mis	25	22.56	.84
Placebo	26	16.00	.54

p value

Comparison	Activity index	Initiates
Epi Inf vs. Epi Mis	.05	.03
Epi Inf vs. Epi Ign	ns	.08
Plac vs. Epi Mis, Ign, or Inf	ns	ns

* Tested by X^2 comparison of the proportion of subjects in each condition initiating new acts.

index" presents summary figures on the extent to which the subject joined in the stooge's activity. This is a weighted index which reflects both the nature of the activities in which the subject engaged and the amount of time he was active. The index was devised by assigning the following weights to the subject's activities: 5—hula hooping; 4—shooting with slingshot; 3—paper airplanes; 2—paper basketballs; 1—doodling; 0—does nothing. Pretest scaling on 15 college students ordered these activities with respect to the degree of euphoria they represented. Arbitrary weights were assigned so that the wilder the activity, the heavier the weight. These weights are multiplied by an estimate of the amount of time the subject spent in each activity and the summed products make up the activity index for each subject. This index may be considered a measure of behavioral euphoria. It should be noted that the same between-condition relationships hold for the two components of this index as for the index itself.

The column labeled "Mean number of acts initiated" presents the data on the extent to which the subject deviates from the stooge's routine and initiates euphoric activities of his own.

On both behavioral indices, we find precisely the same pattern of relationships as those obtained with self-reports. Epi Mis subjects behave somewhat more euphorically than do Epi Ign subjects who in turn behave more euphorically than do Epi Inf subjects. On all measures, then, there is consistent evidence that a subject will take over the stooge's euphoric mood to the extent that he has no other explanation of his bodily state.

Again it should be noted that on these behavioral indices, Epi Ign and Epi Mis subjects are somewhat more euphoric than placebo subjects but not significantly so.

Anger: Self-report. Before presenting data for the anger conditions, one point must be made about the anger manipulation. In the situation devised, anger, if manifested, is most likely to be directed at the experimenter and his annoyingly personal questionnaire. As we subsequently discovered, this was rather unfortunate, for the subjects, who had volunteered for the experiment for extra points on their final exam, simply refused to endanger these points by publicly blowing up, admitting their irritation to the experimenter's face or spoiling the questionnaire. Though as the reader will see, the subjects were quite willing to manifest anger when they were alone with the stooge, they hesitated to do so on material (self-ratings of mood and questionnaire) that the experimenter might see and only after the purposes of the experiment had been revealed were many of these subjects willing to admit to the experimenter that they had been irked or irritated.

This experimentally unfortunate situation pretty much forces us to rely on the behavioral indices derived from observation of the subject's presumably private interaction with the stooge. We do, however, present data on the self-report scales in Table 4. These figures are derived in the same way as the figures presented in Table 2 for the euphoria conditions, that is, the value checked on the irritation scale is subtracted from the value checked on the happiness scale. Though, for the reasons stated above, the absolute magnitude of these figures (all positive) is relatively meaningless, we can, of course, compare condition means within the set of anger conditions. With the happiness-irritation index employed, we should, of course, anticipate precisely the reverse results from those ob-

TABLE 4

SELF-REPORT OF EMOTIONAL STATE IN THE ANGER CONDITIONS

Condition	N	Self-Report scales	Comparison	p
Epi Inf	22	1.91	Epi Inf vs. Epi Ign	.08
Epi Ign	23	1.39	Placebo vs. Epi Ign or Inf	*ns*
Placebo	23	1.63		

tained in the euphoria conditions; that is, the Epi Inf subjects in the anger conditions should again be less susceptible to the stooge's mood and should, therefore, describe themselves as in a somewhat happier frame of mind than subjects in the Epi Ign condition. This is the case; the Epi Inf subjects average 1.91 on the self-report scales while the Epi Ign subjects average 1.39.

Evaluating the effects of the injections, we note again that, as anticipated, Epi Ign subjects are somewhat less happy than Placebo subjects but, once more, this is not a significant difference.

Behavior. The subject's responses to the stooge, during the period when both were filling out their questionnaires, were systematically coded to provide a behavioral index of anger. The coding scheme and the numerical values attached to each of the categories have been described in the methodology section. To arrive at an "Anger index" the numerical value assigned to a subject's responses to the stooge is summed together for the several units of stooge behavior. In the coding scheme used, a positive value to this index indicates that the subject agrees with the stooge's comment and is growing angry. A negative value indicates that the subject either disagrees with the stooge or ignores him.

The relevant data are presented in Table 5. For this analysis, the stooge's routine has been divided into two phases—the first two units of his behavior (the "long" questionnaire and "What did I have for breakfast?") are considered essentially neutral revealing nothing of the stooge's mood; all of the following units are considered "angry" units for they begin with an irritated remark about the "bells" question and end with the stooge's fury as he rips up his questionnaire and stomps out of the room. For the neutral units, agreement or disagreement with the stooge's remarks is, of course, meaningless as an index of mood and we should anticipate no difference between conditions. As can be seen in Table 5, this is the case.

For the angry units, we must, of course, anticipate that subjects in the Epi Ign condition will be angrier than subjects in the Epi Inf condition. This is indeed the case. The Anger index for the Epi Ign condition is positive and large, indicating that these subjects have become angry, while in the Epi Inf condition the Anger index is slightly negative in value indicating that these subjects have failed to catch the stooge's mood at all. It seems clear that providing the subject with an ap-

TABLE 5

BEHAVIORAL INDICATIONS OF EMOTIONAL STATE IN THE ANGER CONDITIONS

Condition	N	Neutral units	Anger units
Epi Inf	22	+0.07	−0.18
Epi Ign	23	+0.30	+2.28
Placebo	22[a]	−0.09	+0.79

Comparison for anger units	p
Epi Inf vs. Epi Ign	<.01
Epi Ign vs. Placebo	<.05
Placebo vs. Epi Inf	*ns*

[a] For one subject in this condition the sound system went dead and the observer could not, of course, code his reactions.

propriate explanation of his bodily state greatly reduces his tendency to interpret his state in terms of the cognitions provided by the stooge's angry behavior.

Finally, on this behavioral index, it can be seen that subjects in the Epi Ign condition are significantly angrier than subjects in the Placebo condition. Behaviorally, at least, the injection of epinephrine appears to have led subjects to an angrier state than comparable subjects who received placebo shots.

Conformation of Data to Theoretical Expectations

Now that the basic data of this study have been presented, let us examine closely the extent to which they conform to theoretical expectations. If our hypotheses are correct and if this experimental design provided a perfect test for these hypotheses, it should be anticipated that in the euphoria conditions the degree of experimentally produced euphoria should vary in the following fashion:

$$\text{Epi Mis} \geqq \text{Epi Ign} > \text{Epi Inf} = \text{Placebo}$$

And in the anger conditions, anger should conform to the following pattern :

$$\text{Epi Ign} > \text{Epi Inf} = \text{Placebo}$$

In both sets of conditions, it is the case that emotional level in the Epi Mis and Epi Ign conditions is considerably greater than that achieved in the corresponding Epi Inf conditions. The results for the Placebo condition, however, are ambiguous for consistently the Placebo subjects fall between the Epi Ign and the Epi Inf subjects. This is a particularly troubling pattern for it makes it impossible to evaluate unequivocally the effects of the state of physiological arousal and indeed raises serious questions about our entire theoretical structure. Though the emotional level is consistently greater in the Epi Mis and Epi Ign conditions than in the Placebo condition, this difference is significant at acceptable probability levels only in the anger conditions.

In order to explore the problem further, let us examine the experimental factors identified earlier, which might have acted to restrain the emotional level in the Epi Ign and Epi Mis conditions. As was pointed out earlier, the ideal test of our first two hypotheses requires an experimental setup in which the subject has flatly no way of evaluating his state of physiological arousal other than by means of the experimentally provided cognitions. Had it been possible to physiologically produce a state of sympathetic activation by means other than injection, one could have approached this experimental ideal more closely than in the present setup. As it stands, however, there is always a reasonable alternative cognition available to the aroused subject—he feels the way he does because of the injection. To the extent that the subject seizes on such an explanation of his bodily state, we should expect that he will be uninfluenced by the stooge. Evidence presented in Table 6 for the anger condition and in Table 7 for the euphoria conditions indicates that this is, indeed, the case.

As mentioned earlier, some of the Epi Ign and Epi Mis subjects in their answers to the open-end questions clearly attributed their physical state to the injection, e.g., "the shot gave me the shivers." In Tables 6 and 7 such subjects are labeled "Self-informed." In Table 6 it can be seen that the self-informed subjects are considerably less angry than are the remaining subjects; indeed, they are not angry at all. With these self-informed subjects eliminated the difference between the Epi Ign and the Placebo

TABLE 6

THE EFFECTS OF ATTRIBUTING BODILY STATE TO THE INJECTION ON ANGER IN THE ANGER EPI IGN CONDITION

Condition	*N*	Anger index	*p*
Self-informed subjects	3	−1.67	*ns*
Others	20	+2.88	*ns*
Self-informed vs. Others			.05

conditions is significant at the .01 level of significance.

Precisely the same pattern is evident in Table 7 for the euphoria conditions. In both the Epi Mis and the Epi Ign conditions, the self-informed subjects have considerably lower activity indices than do the remaining subjects. Eliminating self-informed subjects, comparison of both of these conditions with the Placebo condition yields a difference significant at the .03 level of significance. It should be noted, too, that the self-informed subjects have much the same score on the activity index as do the experimental Epi Inf subjects (Table 3).

It would appear, then, that the experimental procedure of injecting the subjects, by providing an alternative cognition, has, to some extent, obscured the effects of epinephrine. When account is taken of this artifact, the evidence is good that the state of physiological arousal is a necessary component of an emotional experience for when self-informed subjects are removed, epinephrine subjects give consistent indications of greater emotionality than do placebo subjects.

Let us examine next the fact that consistently the emotional level, both reported and behavioral, in Placebo conditions is greater than that in the Epi Inf conditions. Theoretically, of course, it should be expected that the two conditions will be equally low, for by assuming that emotional state is a joint function of a state of physiological arousal and of the appropriateness of a cognition we are, in effect, assuming a multiplicative function, so that if either component is at zero, emotional level is at zero. As noted earlier this expectation should hold if we can be sure that there is no sympathetic activation in the Placebo conditions. This assumption, of course, is completely unrealistic for the injection of placebo does not prevent sympathetic activation. The experimental situations were fairly dramatic and certainly some of the placebo subjects gave indications of physiological arousal. If our general line of reasoning is correct, it should be anticipated that the emotional level of subjects who give indications of sympathetic activity will be greater than that of subjects who do not. The relevant evidence is presented in Tables 8 and 9.

As an index of sympathetic activation we shall use the most direct and unequivocal measure available—change

TABLE 7

THE EFFECTS OF ATTRIBUTING BODILY STATE TO THE INJECTION ON EUPHORIA IN THE EUPHORIA EPI IGN AND EPI MIS CONDITIONS

Epi Ign

	N	Activity Index	*p*
Self-informed subjects	8	11.63	*ns*
Others	17	21.14	*ns*
Self-informed vs. Others			.05

Epi Mis

	N	Activity Index	*p*
Self-informed subjects	5	12.40	*ns*
Others	20	25.10	*ns*
Self-informed vs. Others			.10

TABLE 8

SYMPATHETIC ACTIVATION AND EUPHORIA IN THE EUPHORIA PLACEBO CONDITION

Subjects whose:	*N*	Activity index	*p*
Pulse decreased	14	10.67	*ns*
Pulse increased or remained same	12	23.17	*ns*
Pulse decrease vs. pulse increase or same			.02

in pulse rate. It can be seen in Table 1 that the predominant pattern in the Placebo condition is a decrease in pulse rate. We shall assume, therefore, that those subjects whose pulse increases or remains the same give indications of sympathetic activity while those subjects whose pulse decreases do not. In Table 8, for the euphoria condition, it is immediately clear that subjects who give indications of sympathetic activity are considerably more euphoric than are subjects who show no sympathetic activity. This relationship is, of course, confounded by the fact that euphoric subjects are considerably more active than noneuphoric subjects—a factor which independent of mood could elevate pulse rate. However, no such factor operates in the anger condition where angry subjects are neither more active nor talkative than calm subjects. It can be seen in Table 9 that Placebo subjects who show signs of sympathetic activation give indications of considerably more anger than do subjects who show no such signs. Conforming to expectations, sympathetic activation accompanies an increase in emotional level.

TABLE 9

SYMPATHETIC ACTIVATION AND ANGER IN ANGER PLACEBO CONDITION

Subjects whose:	*N*[a]	Anger index	*p*
Pulse decreased	13	+0.15	*ns*
Pulse increased or remained same	8	+1.69	*ns*
Pusle decrease vs. pulse increase or same			.01

[a] *N* reduced by two cases owing to failure of sound system in one case and experimenter's failure to take pulse in another.

It should be noted, too, that the emotional levels of subjects showing no signs of sympathetic activity are quite comparable to the emotional level of subjects in the parallel Epi Inf conditions (see Tables 3 and 5). The similarity of these sets of scores and their uniformly low level of indicated emotionality would certainly make it appear that both factors are essential to an emotional state. When either the level of sympathetic arousal is low or a completely appropriate cognition is available, the level of emotionality is low.

DISCUSSION

Let us summarize the major findings of this experiment and examine the extent to which they support the propositions offered in the introduction of this paper. It has been suggested, first, that given a state of physiological arousal for which an individual has no explanation, he will label this state in terms of the cognitions available to him. This implies, of course, that by manipulating the cognitions of an individual in such a state we can manipulate his feelings in diverse directions. Experimental results support this proposition for following the injection of epinephrine, those subjects who had no explanation for the bodily state thus produced, gave behavioral and self-report indications that they had been readily manipulable into the disparate feeling states of euphoria and anger.

From this first proposition, it must follow that given a state of physiological arousal for which the individual has a completely satisfactory explanation, he will not label this state in terms of

the alternative cognitions available. Experimental evidence strongly supports this expectation. In those conditions in which subjects were injected with epinephrine and told precisely what they would feel and why, they proved relatively immune to any effects of the manipulated cognitions. In the anger condition, such subjects did not report or show anger; in the euphoria condition, such subjects reported themselves as far less happy than subjects with an identical bodily state but no adequate knowledge of why they felt the way they did.

Finally, it has been suggested that given constant cognitive circumstances, an individual will react emotionally only to the extent that he experiences a state of physiological arousal. Without taking account of experimental artifacts, the evidence in support of this proposition is consistent but tentative. When the effects of "self-informing" tendencies in epinephrine subjects and of "self-arousing" tendencies in placebo subjects are partialed out, the evidence strongly supports the proposition.

The pattern of data, then, falls neatly in line with theoretical expectations. However, the fact that we were forced, to some extent, to rely on internal analyses in order to partial out the effects of experimental artifacts inevitably makes our conclusions somewhat tentative. In order to further test these propositions on the interaction of cognitive and physiological determinants of emotional state, a series of additional experiments, published elsewhere, was designed to rule out or overcome the operation of these artifacts. In the first of these, Schachter and Wheeler (1962) extended the range of manipulated sympathetic activation by employing three experimental groups—epinephrine, placebo, and a group injected with the sympatholytic agent, chlorpromazine. Laughter at a slapstick movie was the dependent variable and the evidence is good that amusement is a direct function of manipulated sympathetic activation.

In order to make the epinephrine-placebo comparison under conditions which would rule out the operation of any self-informing tendency, two experiments were conducted on rats. In one of these Singer (1961) demonstrated that under fear inducing conditions, manipulated by the simultaneous presentation of a loud bell, a buzzer, and a bright flashing light, rats injected with epinephrine were considerably more frightened than rats injected with a placebo. Epinephrine-injected rats defecated, urinated, and trembled more than did placebo-injected rats. In nonfear control conditions, there were no differences between epinephrine and placebo groups, neither group giving any indication of fear. In another study, Latané and Schachter (1962) demonstrated that rats injected with epinephrine were notably more capable of avoidance learning than were rats injected with a placebo. Using a modified Miller-Mowrer shuttlebox, these investigators found that during an experimental period involving 200 massed trials, 15 rats injected with epinephrine avoided shock an average of 101.2 trials while 15 placebo-injected rats averaged only 37.3 avoidances.

Taken together, this body of studies does give strong support to the propositions which generated these experimental tests. Given a state of sympathetic activation, for which no immediately appropriate explanation is available, human subjects can be readily manipulated into states of euphoria, anger, and amusement. Varying the intensity of sympathetic activation serves to vary the intensity of a variety of emotional states in both rats and human subjects.

Let us examine the implications of these findings and of this line of thought for problems in the general area of the physiology of the emotions. We have noted in the introduction that the numerous studies on physiological differentiators of emotional states have, viewed en masse, yielded quite inconclusive results. Most, though not all, of these studies have indicated no differences among the various emotional states. Since as human beings, rather than as scientists, we have no difficulty identifying, labeling, and distinguishing among our feelings, the results of these studies have long seemed rather puzzling and paradoxical. Perhaps because of this, there has been a persistent tendency to discount such results as due to ignorance or methodological inadequacy and to pay far more attention to the very few studies which demonstrate *some* sort of physiological differences among emotional states than to the very many studies which indicate no differences at all. It is conceivable, however, that these results should be taken at face value and that emotional states may, indeed, be generally characterized by a high level of sympathetic activation with few if any physiological distinguishers among the many emotional states. If this is correct, the findings of the present study may help to resolve the problem. Obviously this study does *not* rule out the possibility of physiological differences among the emotional states. It is the case, however, that given precisely the same state of epinephrine-induced sympathetic activation, we have, by means of cognitive manipulations, been able to produce in our subjects the very disparate states of euphoria and anger. It may indeed be the case that cognitive factors are major determiners of the emotional labels we apply to a common state of sympathetic arousal.

Let us ask next whether our results are specific to the state of sympathetic activation or if they are generalizable to other states of physiological arousal. It is clear that from our experiments proper, it is impossible to answer the question for our studies have been concerned largely with the effects of an epinephrine created state of sympathetic arousal. We would suggest, however, that our conclusions are generalizable to almost any pronounced internal state for which no appropriate explanation is available. This suggestion receives some support from the experiences of Nowlis and Nowlis (1956) in their program of research on the effects of drugs on mood. In their work the Nowlises typically administer a drug to groups of four subjects who are physically in one another's presence and free to interact. The Nowlises describe some of their results with these groups as follows:

> At first we used the same drug for all 4 men. In those sessions seconal, when compared with placebo, increased the checking of such words as expansive, forceful, courageous, daring, elated, and impulsive. In our first statistical analysis we were confronted with the stubborn fact that when the same drug is given to all 4 men in a group, the N that has to be entered into the analysis is 1, not 4. This increases the cost of an already expensive experiment by a considerable factor, but it cannot be denied that the effects of these drugs may be and often are quite contagious. Our first attempted solution was to run tests on groups in which each man had a different drug during the same session, such as 1 on seconal, 1 on benzedrine, 1 on dramamine, and 1 on placebo. What does seconal do? Cooped up with, say, the egotistical benzedrine partner, the withdrawn, indifferent dramimine partner, and the slightly bored lactose man, the seconal subject reports that he is distractible, dizzy, drifting, glum, defiant, languid, sluggish, discouraged, dull, gloomy, lazy, and slow! This is not the report of mood that we got when all 4 men were on seconal. It thus appears that the moods of the partners do definitely influence the effect of seconal (p. 350).

It is not completely clear from this description whether this "contagion"

of mood is more marked in drug than in placebo groups, but should this be the case, these results would certainly support the suggestion that our findings are generalizable to internal states other than that produced by an injection of epinephrine.

Finally, let us consider the implications of our formulation and data for alternative conceptualizations of emotion. Perhaps the most popular current conception of emotion is in terms of "activation theory" in the sense employed by Lindsley (1951) and Woodworth and Schlosberg (1958). As we understand this theory, it suggests that emotional states should be considered as at one end of a continuum of activation which is defined in terms of degree of autonomic arousal and of electroencephalographic measures of activation. The results of the experiment described in this paper do, of course, suggest that such a formulation is not completely adequate. It is possible to have very high degrees of activation without a subject either appearing to be or describing himself as "emotional." Cognitive factors appear to be indispensable elements in any formulation of emotion.

SUMMARY

It is suggested that emotional states may be considered a function of a state of physiological arousal and of a cognition appropriate to this state of arousal. From this follows these propositions:

1. Given a state of physiological arousal for which an individual has no immediate explanation, he will label this state and describe his feelings in terms of the cognitions available to him. To the extent that cognitive factors are potent determiners of emotional states, it should be anticipated that precisely the same state of physiological arousal could be labeled "joy" or "fury" or "jealousy" or any of a great diversity of emotional labels depending on the cognitive aspects of the situation.

2. Given a state of physiological arousal for which an individual has a completely appropriate explanation, no evaluative needs will arise and the individual is unlikely to label his feelings in terms of the alternative cognitions available.

3. Given the same cognitive circumstances, the individual will react emotionally or describe his feelings as emotions only to the extent that he experiences a state of physiological arousal.

An experiment is described which, together with the results of other studies, supports these propositions.

REFERENCES

AX, A. F. Physiological differentiation of emotional states. *Psychosom. Med.*, 1953, **15**, 433–442.

CANNON, W. B. *Bodily changes in pain, hunger, fear and rage.* (2nd ed.) New York: Appleton, 1929.

CANTRIL, H., & HUNT, W. A. Emotional effects produced by the injection of adrenalin. *Amer. J. Psychol.*, 1932, **44**, 300–307.

FESTINGER, L. A theory of social comparison processes. *Hum. Relat.*, 1954, **7**, 114–140.

HUNT, J. MCV., COLE, M. W., & REIS, E. E. Situational cues distinguishing anger, fear, and sorrow. *Amer. J. Psychol.*, 1958, **71**, 136–151.

JAMES, W. *The principles of psychology.* New York: Holt, 1890.

LANDIS, C., & HUNT, W. A. Adrenalin and emotion. *Psychol. Rev.*, 1932, **39**, 467–485.

LATANÉ, B., & SCHACHTER, S. Adrenalin and avoidance learning. *J. comp. physiol. Psychol.*, 1962, **65**, 369–372.

LINDSLEY, D. B. Emotion. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 473–516.

MARAÑON, G. Contribution à l'étude de l'action émotive de l'adrénaline. *Rev. Francaise Endocrinol.*, 1924, **2**, 301–325.

NOWLIS, V., & NOWLIS, H. H. The description and analysis of mood. *Ann. N. Y. Acad. Sci.*, 1956, **65**, 345–355.

RUCKMICK, C. A. The psychology of feeling *and emotion.* New York: McGraw-Hill, 1936.

SCHACHTER, J. Pain, fear, and anger in hypertensives and normotensives: A psychophysiologic study. *Psychosom. Med.*, 1957, **19**, 17–29.

SCHACHTER, S. *The psychology of affiliation.* Stanford, Calif.: Stanford Univer. Press, 1959.

SCHACHTER, S., & WHEELER, L. Epinephrine, chlorpromazine, and amusement. *J. abnorm. soc. Psychol.*, 1962, **65**, 121–128.

SINGER, J. E. The effects of epinephrine, chlorpromazine and dibenzyline upon the fright responses of rats under stress and non-stress conditions. Unpublished doctoral dissertation, University of Minnesota, 1961.

WOLF, S., & WOLFF, H. G. *Human gastric function.* New York: Oxford Univer. Press, 1947.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology.* New York: Holt, 1958.

WRIGHTSMAN, L. S. Effects of waiting with others on changes in level of felt anxiety. *J. abnorm. soc. Psychol.*, 1960, **61**, 216–222.

(Received February 17, 1961)

Psychological Review
1962, Vol. 69, No. 5, 400–414

ON THE QUANTIFICATION OF MILLER'S CONFLICT THEORY[1]

NORMAN H. ANDERSON

University of California, Los Angeles

This paper elucidates a mathematical model of potential usefulness in N. E. Miller's theory of conflict. Miller has discussed his theory most extensively in his provocative 1959 article and the present model is based fairly directly on that discussion. For simplicity of exposition, the model will be discussed in a rather restricted context. Only two particular experimental designs for a specific experimental task involving approach-avoidance conflict will be considered. However, the basic ideas and the techniques used have a considerably greater range of applicability.

In the experimental task to be considered, the human subject taps a key once upon presentation of a signal which is set to repeat at timed intervals. A goal, which may be represented by illuminating one of a row of lamps, moves one step closer to the subject at each tap of the key. When the last tap is made, the goal moves the last step, and the subject receives reward and punishment (e.g., coin and shock), either simultaneously or, in certain cases, on different trials. The dependent variable is response speed (reciprocal latency) as a function of distance from the goal. Each trial thus yields essentially a complete curve as the datum.

Provision for an overt retreat response could also be made and would possibly be useful in studying choice point vacillation. However, it will be assumed that only an approach or nonapproach response is permitted and, moreover, that the reward and punishment conditions are such that the subject reaches the goal on each trial. This latter restriction simplifies the presentation by avoiding, in particular, the more detailed statistical discussion needed when some response curves are incomplete.

The task described breaks up the goal response into a sequence of subresponses which has a number of advantages. The successive points on the response curve are more nearly physiologically independent than if a free continuous response were used. Since the goal moves, rather than the subject, the motor aspect of the successive subresponses should remain the same. In addition, various reinforcement schedules may be programed over the last several steps to the goal. Finally, the use of discrete subresponses provides simultaneous information on response probability and response strength which should be useful theoretically (Estes, 1950; Mueller, 1950; Spence, 1954).

MODEL

Assumptions

The model rests on the following assumptions:

A_1. The approach and avoidance gradients exist and are independent.

A_2. Response speed is proportional to the difference between the approach and avoidance gradients. Since the constant of proportionality represents only the choice of units on the meas-

[1] A preliminary version of this paper was presented at the first annual meeting of the Psychonomic Society, September 1960, in Chicago.

urement scales, it may conveniently be taken equal to unity. The assumption may then be written:

$$S(x) = F(x) - G(x),$$

where x is distance from the goal, $F(x)$ is the approach gradient, $G(x)$ is the avoidance gradient, and $S(x)$ is response speed.

A_3. The approach and avoidance gradients can be written in the following form:

$$F(x) = af(x) + b + c,$$
$$G(x) = dg(x) + e + f,$$

where a, b are parameters depending on reward; d, e are parameters depending on punishment; c, f are constants in the given experimental situation; $f(x)$, $g(x)$ are completely arbitrary functions. The functions, $f(x)$ and $g(x)$, will be called the shape functions of the approach and avoidance gradients, respectively. The six parameters appearing in A_3 are assumed to vary with individuals and with trials. However it will be assumed here that no such variation occurs in the shape functions. This latter assumption will be discussed further under Application 2.

Of the three assumptions of the model, A_1 is the most stringent. The statement that the approach and avoidance tendencies exist is the main assertion about the psychological processes underlying the behavior. The statement that the gradients are independent is a brief way of saying that reward affects only the approach gradient and that punishment affects only the avoidance gradient. Of these two independence propositions, the first seems not unreasonable but the second should perhaps be viewed with greater reserve. Nevertheless, Miller does make these assumptions, both explicitly in Figures 7 and 9 of the cited reference (Miller, 1959), and implicitly in his procedure of comparing separately obtained approach and avoidance response curves. These assumptions will accordingly be used here. A somewhat weaker independence requirement will be discussed below.

The proportionality assumption, A_2, is more stringent than that made by Miller (1959) who requires only that the response measure be some monotone function of the difference between the gradients. It is, however, not entirely out of line with Hull-Spence theory (e.g., Logan, 1952; Spence, 1956) if the gradients are interpreted as reaction potentials. In any case, it will be seen below that only monotonicity is required if an appropriate experimental design is used.

Assumption A_3 is fairly weak since no restriction is placed on $f(x)$ and $g(x)$; these functions are entirely arbitrary and need not even be estimated for the proposed tests of the model. The given representation of the approach gradient assumes that the effect of reward is partly multiplicative and partly additive, as represented by the parameters a and b, respectively, and an analogous statement holds for punishment. The use of six parameters may seem peculiarly generous but the first test of the model will be completely parameter free, and the second test requires evaluation of only two parameters.

Before proceeding to the derivations, a few remarks on averaging are required since biases may result from differences between individuals or trials (Anderson, 1959, 1960, p. 85). For the present case, consider a given set of subjects and trials. Because of individual differences, subjects treated alike will in general have different gradients. Moreover, gradients will vary with trials because of learning,

random fluctuation, or adaptation effects of various kinds. To the given set of subjects and trials therefore, corresponds a set or distribution of gradients of either type. Assumption A_1 implies that experimental conditions with say, the same punishment but different rewards will have the same distribution of avoidance gradients. This conclusion follows from A_1 under the present restriction that the subjects reach the goal on each trial, thus receiving both reward and punishment on each trial. Given this restriction, the assumption that reward and punishment act independently implies that the experimental conditions provide equivalent punishment treatment, and so yield identical distributions of avoidance gradients. In the same way, it is seen that experimental conditions with the same reward but different punishment have the same mean approach gradient. These two results are used to justify the use of average data in the following applications of the model.

Application 1

This application is based on a 2×2 design with two values each of reward and punishment. Consider a given subject receiving Reward i and Punishment j, and let $F_i(x)$ be the approach gradient, $G_j(x)$ the avoidance gradient, and $S_{ij}(x)$ the speed function for this subject on a given trial. Then by A_2:

$$S_{ij}(x) = F_i(x) - G_j(x). \quad [1]$$

This result holds for a given subject on a given trial. If a set of subjects and trials is considered, there will be a corresponding set or distribution of gradients. As shown above, the mean approach gradient, averaged over this distribution, will be the same for the two experimental conditions which have the same reward. Similarly, the mean avoidance gradient will depend only on the punishment involved, not on the reward. Since the mean of the difference on the right side of Equation 1 is the difference of the means, it follows that Equation 1 holds on the average over the set of subjects and trials under consideration. The quantities appearing in Equation 1 will accordingly be interpreted as means. When the four theoretical mean speed functions of Equation 1 are placed in the 2×2 table of the design, the following array is obtained:

$$\begin{matrix} F_1(x) - G_1(x) & F_1(x) - G_2(x) \\ F_2(x) - G_1(x) & F_2(x) - G_2(x). \end{matrix} \quad [2]$$

Inspection of this array shows that the difference between the two entries in the first row is equal to the difference between the two entries in the second row. This means, in analysis of variance terms, that the interaction of the row (reward) and column (punishment) variables is zero at each value of x. The model thus implies that the Reward $\times$ Punishment, and the Reward $\times$ Punishment $\times$ Distance interactions are zero in principle, and hence nonsignificant in practice. Consequently, the ordinary analysis of variance may be applied directly to the obtained speed curves in order to test the model.

This application of the model requires only A_1 and A_2 and is completely parameter free. Analogous results hold, of course, for any factorial design with reward and punishment as two of the factors.

Application 2

Consider now a situation with fixed punishment and g conditions of reward. The speed curve on a given trial for a given subject under Reward

i may be written:

$$S_i(x) = a_i f(x) + b_i + c - G(x) \qquad i = 1, 2, \cdots, g. \quad [3]$$

If Equation 3 is to be useful, it must be averaged over the distribution of gradients corresponding to a given set of subjects and trials. Now the average of the sum of terms on the right side of the equation is the sum of the average terms. As noted above, the mean $G(x)$ is the same in all reward conditions, and so also is the mean value of c. Since $f(x)$ was taken as fixed, the mean of the product, $a_i f(x)$, is just $f(x)$ times the mean a_i. Hence Equation 3 is valid if each term is interpreted as a mean and this will be done.

A rough test of the model for this application may be obtained by using Equation 3 to get expressions for, say, $S_2(x) - S_1(x)$, and $S_3(x) - S_1(x)$. These expressions may fairly easily be combined to show that the one difference is a simple linear function of the other difference:

$$S_3(x) - S_1(x) = A[S_2(x) - S_1(x)] + B, \quad [4]$$

where A and B are constants whose exact expressions in terms of the a_i and b_i are not of importance. The values of A and B could be estimated by using selected values of x, or by least squares. An assessment of goodness of fit could then be made by eye and this procedure, crude though it is, might be useful in many cases.

A rigorous statistical test for Application 2 may be obtained using the multivariate analysis of variance in the form of Fisher's (1938) test of coplanarity. For suppose Equation 3 is correct, and that the speed gradients are measured at, say, p values of x. Then it can readily be shown that the $p \times p$ between conditions sums of products matrix has Rank 2 ($g \geq 3$). The null hypothesis to be tested is then, in Rao's (1952) notation, that the smallest $p - 2$ latent roots of $B - \lambda W$ are zero, where B and W are the between and within conditions sum of products matrices, respectively. Rao gives numerical examples (pp. 364–375). The tests are approximate but evidently quite adequate in respect of significance level. The results of Marriott (1956) might be expected to yield somewhat more powerful tests. Unfortunately both procedures are somewhat tedious with a desk calculator.

Application 2 requires only half of A_1, namely, that reward does not affect the avoidance gradient. Also only half of A_3 is used since no restriction is placed on $G(x)$. The treatment of the case of fixed reward and variable punishment can be done in essentially the same way.

Since Application 2 uses the approximate approach gradient representation of A_3 the results are not exact, in contrast to those of the first application. It is interesting to note that the same results are obtained if a_i is assumed fixed within each reward condition and $f(x)$ is allowed to vary over trials and subjects. It would still be required that the mean $f(x)$ be the same in each condition, but the individual subjects could have sharply different approach gradient shape functions.

If both a_i and $f(x)$ were to be considered as variable over subjects and trials, each individual approach gradient would be represented exactly. However, if both a_i and $f(x)$ are variable, the mean of their product is the product of their means, plus the covariance of the two factors. In this case, Equation 3, interpreted in terms of means, would be biased because this covariance term is not included. Obversely, the covariance

term indexes the adequacy of the approximation of the gradient representation of A_3. The bias in Equation 3 tends to cancel in the tests of goodness of fit and is not thought to be serious. However, it should be noted that the bias may be reduced by reducing the variability between trials and between subjects. Trial variability may be reduced by restricting attention to asymptotic data. The subject variability may be reduced by considering homogeneous blocks of subjects including, in the limit, single subjects.

Comments

If the model fails the experimental tests of the sort illustrated in the above two applications, one or more of the assumptions must be at fault.

The proportionality assumption, A_2, although presumably a reasonable first approximation, may well not be sufficiently exact for the quantitative tests under consideration. However, if the response measure is not proportional to the difference between the gradients, it should in any case be some strictly monotone function of this difference. From this it follows, by taking inverse functions, that there is some monotone function of the response measure which *is* proportional to the difference between the two gradients. Thus, if A_2 alone is at fault, it may in principle be rectified by employing this latter monotone function of the response measure in place of the original response measure. The only problem is whether this scaling of the response measure may be accomplished in practice. For the cited applications, the scaling presents no particular difficulty and the details are given in a subsequent section.

It thus follows that, if no scaling of the data satisfies the model, one of the other assumptions must be incorrect. In this case, since A_3 is quite weak and is used only in Application 2, A_1 would become suspect.

Assumption A_1 might be incorrect for two different reasons. First, the approach and avoidance gradients might give an inadequate representation of the processes underlying the behavior. Indeed, the present development, in contrast to that of Bower (1959) and Spence (1960, Ch. 22), does not give an explicit account of choice point vacillation. Further investigation of this aspect of the behavior would probably be quite useful as Miller (1959) has observed, and would be expected to guide any necessary revisions of the two gradient representation.

The second reason for rejection of A_1 would be nonindependence of the approach and avoidance gradients. If in fact reward and punishment affect both gradients, it will probably be necessary to include the learning processes explicitly in the model. This problem has not been investigated in any detail since it seems preferable to base the initial tests of the model on the simplest assumptions.

If the model is supported by the experimental tests, one's confidence in the assumptions would be increased, but it would still be useful to ask how much had been learned. Success in prediction gives only an indirect confirmation of the premises which yielded the prediction. Thus successful prediction would not prove A_1 because alternative assumptions might imply the same results as, indeed, the following example shows.

Consider Application 2 with reward and punishment given on separate trials. Suppose that the avoidance gradient has the same shape function as the approach gradient. Equation

3 may then be written:

$$S_i(x) = (a_i - d_i)f(x) + (b_i - e_i) + (c - f).$$

Suppose also that there is a decrement in the avoidance gradient on rewarded trials which results simply from the absence of punishment and so is independent of reward magnitude. This decrement would be represented by decreases in d_i and e_i in the equation above. Inspection of the equation shows that it is impossible to distinguish decreases in d_i and e_i from corresponding increases in a_i and b_i. It would thus be impossible to distinguish a simultaneous increase in $F(x)$ and decrease in $G(x)$ from a larger increase in $F(x)$ and no change in $G(x)$. Consequently, if it is also assumed that any effect of punishment on the reward gradient is independent of amount of reward, then the results derived in Application 2 hold also in the present case.

The example of the preceding paragraph, which seems to be a fairly reasonable possibility, indicates that the model will hold under a wider range of assumptions than those actually used in the derivation. This extra generality is a mixed blessing since, except for certain practical situations, prediction of behavior is only a means to the end of understanding behavior. However, the model does discriminate between variants of A_1 which are of interest. Moreover, it does not seem likely that the model predictions would be verified unless the measuring scale for the dependent variable were correct.

The estimation and interpretation of the shape functions and of the reward and punishment parameters also require discussion. The representation of the gradients used in A_2 is quite flexible and should give a good fit to any particular set of gradients. However, it is a representation of convenience, one which dissects the gradients in a mathematically powerful but possibly unpsychological way. Consequently, the reward and punishment parameters do not necessarily have any direct psychological interpretation. Whether or not they do cannot be determined without further empirical or theoretical analysis. Since the tests of the model do not require the estimation of these parameters, this problem is not considered here. Some comments on the interpretation of the parameters are given in a subsequent section.

Matters are somewhat clearer as regards the shape functions. If the model survives the test of Application 2, then Equation 3 may be used to estimate the shape function of the approach gradient. This shape function is only unique up to a multiplicative factor and an additive term so that it might perhaps be desirable to reduce to a standard form by setting $f(0) = 1$, $f(\infty) = 0$. How generally useful the resultant estimate would be is an open question. As Miller (1959) indicates, the gradients will be strongly influenced by the stimulus properties of the situation. There would thus be no reason to expect more than a rough similarity between shape functions from different experiments unless the procedures were carefully standardized. Within a given situation, however, estimation of the shape functions would be expected to be relevant to Miller's hypothesis that the avoidance gradient has a steeper slope than the approach gradient.

Another problem deserving comment arises because the present applications of the model are based on comparisons of different experimental conditions in which either the ap-

proach gradient or the avoidance gradient is the same in two or more conditions. If the present restriction that the subjects reach the goal on every trial is not met, the required comparability may not obtain. Thus subjects receiving a greater punishment might reach the goal less often than subjects receiving a lesser punishment. If so, the number of rewards would not be the same in the two conditions, the approach gradients would not be the same, and the above applications need not hold. In situations of this sort, there may be some justification for applying the model to the asymptotic data. For, by A_1, the same asymptotic reward gradients should be obtained, the lower rate of reward in the condition with greater punishment affecting only the rate at which the asymptote is reached.

In this connection, it is worth mentioning a potentially useful experimental procedure. The subject is run under a given reward and punishment until the behavior has stabilized and then told that on the next trial the reward will be increased. By A_1, this announcement should affect only the approach gradient. The avoidance gradients would be the same on the trial preceding as on the trial following the reward so that the data from these two trials would be comparable in the desired way.

Finally it should be noted that the model may be more generally applicable than has been indicated. It might be possible, depending on the experimental setup, to consider the two gradients as approach gradients with respect to two incompatible goals. In particular, some of the results of Application 1 hold even when there is but one step from start to goal. These results would be potentially applicable to situations involving two competing response tendencies.

LEARNING

Explicit incorporation of the learning processes into the model appears to present some difficulties. This section outlines one possible approach which should help set out the problems involved. The discussion is not essential to the rest of the paper and only a very brief sketch will be given.

For illustrative purposes, it is assumed that reward is given on a random proportion, π, of the trials, punishment being given on the remaining $1 - \pi$ of the trials. In the learning model to be considered, it is assumed that each gradient changes on each trial by a constant fraction of the possible change. This proportional change assumption is fairly common (Anderson, 1961; Bush & Mosteller, 1955; Estes, 1950; Hull, 1943) but it is here being applied to the whole gradient. Evidently it is necessary to allow the possibility of different constant fractions for the two gradients and for the two types of trials.

If reward is given on Trial n, the gradients on Trial $n + 1$ are:

$$F_{n+1}(x) = F_n(x) + \theta_1[F_U(x) - F_n(x)], \quad [5]$$

$$G_{n+1}(x) = G_n(x) + \theta_2[G_L(x) - G_n(x)].$$

If punishment is given on Trial n, the gradients on Trial $n + 1$ are:

$$F_{n+1}(x) = F_n(x) + \theta_1'[F_L(x) - F_n(x)], \quad [6]$$

$$G_{n+1}(x) = G_n(x) + \theta_2'[G_U(x) - G_n(x)].$$

Here θ_1, θ_2, θ'_1, θ'_2 are the four learning rates; subscripts n and $n + 1$ index the gradients on Trials n and $n + 1$; subscripts U and L index the upper and lower limit gradients, quantities

analogous to the limit points of the linear operators of Bush and Mosteller (1955). The quantities in brackets thus denote the possible change on Trial n, and the equations state simply that the actual change is a constant fraction of the possible change.

Although the learning model is conceptually simple, its complete analysis would involve the estimation of four learning rates and four limit gradients. It seems reasonable to suppose that two of the limit gradients, $F_L(x)$ and $G_L(x)$, are identically zero, and this will be done in what follows. Further restrictions on the parameters seem problematical so that the estimation problem would be difficult. However, analysis of the sequential dependencies, that is, of the trial-to-trial increments and decrements in the response measure, does yield some information which may be useful.

On the basis of the above formulation, the speed increment on rewarded trials, denoted by $C_R(x)$, can be shown to be:

$$C_R(x) = \theta_1[F_U(x) - F_n(x)] + \theta_2 G_n(x).$$

On punished trials, the speed decrement is:

$$C_P(x) = \theta'_1 F_n(x) + \theta'_2[G_U(x) - G_n(x)].$$

The sum of the increment and decrement is thus:

$$C_R(x) + C_P(x) = \theta_1 F_U(x) + \theta'_2 G_U(x) + (\theta'_1 - \theta_1)F_n(x) - (\theta'_2 - \theta_2)G_n(x). \quad [7]$$

The left side of the equation may be obtained from the data. For a given set of trials, $C_R(x)$ would be estimated by the speed increment averaged over all rewarded trials; similarly $C_P(x)$ would be estimated by the speed decrement averaged over all punished trials. $F_n(x)$ and $G_n(x)$ would correspondingly be replaced by their theoretical averages over the given set of trials.

Now $F_n(x)$ and $G_n(x)$ depend on the value of π; however, the other quantities on the right side of Equation 7 are independent of π. If reward and punishment had equivalent effects so that $\theta'_1 = \theta_1$, and $\theta'_2 = \theta_2$, then the last two terms on the right side vanish. In this case, the model implies that $C_R(x) + C_P(x)$ has the same values for different groups run under different values of π. In this case, therefore, the sequential dependency provides an exact test of the model. Since the last two terms of Equation 7 involve learning rate differences, the sequential dependency should be approximately independent of π as long as the effects of reward and punishment are not too disparate.

If punishment is fairly strong compared to reward, it would be expected that both $\theta'_1 > \theta_1$ and $\theta'_2 > \theta_2$. Then, since $F_n(x)$ is an increasing function of π, and $G_n(x)$ is a decreasing function of π, the sequential dependency should show modest increases as π is increased. Conversely, if punishment is fairly weak compared to reward, it would be expected that $\theta'_1 < \theta_1$, and $\theta'_2 < \theta_2$. The sequential dependency would then show modest decreases as π is increased.

Related results may be obtained from the higher order sequential dependencies, and one of these deserves particular mention. Let $C_{PR}(x)$ denote the change in response caused by a pair of successive trials, the first of which is punished and the second rewarded. Similarly, let $C_{RP}(x)$ be the speed change caused by a rewarded and a punished trial given in that order. The differences between these quantities is always positive and

indeed:

$$C_{PR}(x) - C_{RP}(x) = \theta_1\theta'_1F_U(x) + \theta_2\theta'_2G_U(x). \quad [8]$$

The difference on the left side of this equation is to be estimated from the data in the same way as mentioned above. The equation asserts that this difference is independent of π. This result affords a rigorous nonparametric test of goodness of fit. It should also be noted that the difference of Equation 8 varies directly with the learning rates thus yielding a qualitative check of the model.

The above results are less complete than would be desired but they do permit an assessment of the model. Although in principle the sequential dependencies provide the information for a complete solution to the parameter estimation problem (Anderson, 1959), practical difficulties are great. Given two types of gradient and two types of reinforcing stimuli, however, there are necessarily a considerable number of quantities to be estimated. Consequently, it may be useful to consider the possibility of a representation of the conflict situation which is mathematically more tractable even if possibly less realistic. It should be noted, however, that the trial-to-trial increments and decrements in response strength are the basic quantities in the study of learning, and their investigation should prove valuable under any circumstances.

SCALING THE RESPONSE MEASURE

This section shows how the scaling of the response measure mentioned above may be accomplished. The method is general in that it applies to any continuous response measure. However, the scaling equations are fairly specific to the particular experimental design employed so that a general presentation is not possible. Consequently, only a brief sketch indicating how the method would be used in the above two applications will be given. Statistical details, such as the loss of *df* in estimating the auxiliary scaling parameters, will not be considered.

The basic idea of the present method is to use the theoretical equations directly in the scaling procedure. This approach may be more generally applicable. To the author, it seems considerably more attractive than Hull's (Hull, Felsinger, Gladstone, & Yamaguchi, 1947) quantification procedure because the psychophysical-psychometric assumptions on which the latter is based are not closely related to the form of the theoretical behavior laws proposed for investigation.

If scaling is possible in principle, then, as noted above, there is some monotone function, $\mathfrak{M}$, of the observed response measure, $S(x)$, which does satisfy the proportionality assumption, A_2. Expanding $\mathfrak{M}[S(x)]$ in a Maclaurin series yields:

$$\mathfrak{M}[S(x)] = S(x) + \alpha_2S^2(x) + \alpha_3S^3(x) + \cdots. \quad [9]$$

The constant term in the series expansion does not appear since it is required that $\mathfrak{M}(0) = 0$. Moreover, it is no restriction on present purposes to take the coefficient, α_1, of $S(x)$ to be unity and this has been done. The number of terms of the series which would be required depends on how well the original scale approximates the final scale; only the first three terms of the series will be used here since it would seem that that should be adequate for most work.

Consider first Application 1 and for convenience of notation let:

$$P_n(x) = S_{11}{}^n(x) - S_{12}{}^n(x) - S_{21}{}^n(x) + S_{22}{}^n(x), \qquad n = 1, 2, \cdots.$$

The assumption that $S(x)$ is proportional to the difference between the gradients is equivalent to the requirement that the Reward × Punishment interactions of Equation 2 be zero. These interactions will be zero when the difference between the entries in the first row equals the difference between the entries in the second row for each value of x. In present notation, this condition means concisely that $P_1(x) = 0$ for each x.

If $\mathfrak{M}[S(x)]$, rather than $S(x)$ itself, is to be proportional to the difference between the gradients, a condition analogous to $P_1(x) = 0$ must hold. If $S(x)$ in Equation 2 is replaced by $\mathfrak{M}[S(x)]$ as represented in the series expansion of Equation 9, this condition is seen to be:

$$P_1(x) + \alpha_2 P_2(x) + \alpha_3 P_3(x) = 0.$$

Scaling is thus equivalent to choosing α_2 and α_3 so that this last equation is true for each x. In practice, of course, this would not be possible because of sampling error. Consequently, α_2 and α_3 are chosen so as to minimize the interaction, using the method of least squares (e.g., Lewis, 1960). The sum of squared deviations is $\Sigma_x[P_1(x) + \alpha_2 P_2(x) + \alpha_3 P_3(x)]^2$. This sum is to be differentiated with respect to α_2 and α_3 and these derivatives set equal to zero. The following two linear equations result:

$$\Sigma_x P_1(x)P_2(x) + \alpha_2 \Sigma_x P_2^2(x) + \alpha_3 \Sigma_x P_2(x)P_3(x) = 0,$$

$$\Sigma_x P_1(x)P_3(x) + \alpha_2 \Sigma_x P_2(x)P_3(x) + \alpha_3 \Sigma_x P_3^2(x) = 0.$$

Since the values of the $P_n(x)$ are obtainable from the raw data, these equations are easily solved for α_2 and α_3. With these values of α_2 and α_3, the observed responses would be scaled using Equation 9 and the analysis of variance applied directly to the scaled data. If the interaction remains significant in the scaled data, then the scaled response measure does not satisfy the proportionality assumption. If additional terms of the series expansion of Equation 9 were used, say $N + 1$ altogether, the scaling method would yield the following N linear equations in N unknowns:

$$\Sigma_x \Sigma_s \alpha_s P_s(x) P_t(x) = 0, \qquad t = 2, 3, \cdots, N + 1.$$

The scaling technique for Application 2 follows similar lines. Let:

$$R_n(x) = S_3{}^n(x) - AS_2{}^n(x) + (A - 1)S_1{}^n(x), \quad n = 1, 2, \cdots.$$

From Equation 4, the requirement of the model in terms of the original response measure is that $R_1(x) - B = 0$. In terms of the scaled measure, this requirement becomes:

$$R_1(x) + \alpha_2 R_2(x) + \alpha_3 R_3(x) - B = 0.$$

It is thus required to minimize $\Sigma_x[R_1(x) + \alpha_2 R_2(x) + \alpha_3 R_3(x) - B]^2$. This minimization may be accomplished by using least squares to obtain four equations in four unknowns, α_2, α_3, A, B. The equations, which are not given here, are unfortunately mildly nonlinear in the unknowns so that a process of successive approximation would be required for their solution.

Scales obtained from the above procedure are subject to a bias since the mean of the transformed scores will not in general be exactly equal to the transform of the mean score. The amount of bias would be expected to be negligible unless $\mathfrak{M}$ is considerably nonlinear. If $\mathfrak{M}$ is badly nonlinear, it would probably be advisable to iterate the scaling procedure in order to converge on an unbiased scale.

If the $\mathfrak{M}$ obtained from the scaling procedure is nonmonotone, then one must of course reject the possibility of

a theoretically adequate scaling, and hence also the basic theory. However, it is apparent that the scaling technique may be rather too obliging in yielding too readily, in a given experimental situation, a transformation of the observed response measure which satisfies the model. Nevertheless, there is a requirement of generality implicit in the notion of a measurement scale; it must be applicable to a somewhat wider class of situations than that on which the scaling is based. Essentially this problem has been discussed by Seward (1955) as a requirement of constancy of the intervening variable. It suffices here to note that this requirement of generality also means that the scaling may be done in a convenient situation. The resultant scale may then be used to test the model in situations in which the solution of the scaling equations would be difficult or impossible.

In this connection, it should be observed that the scaling procedure itself may be employed in such a way as to give a consistency check. If the distance variable is divided into subranges, the scaling procedure may be applied to the data of each separately. Theoretically, the same scale should be obtained in each subrange so that comparison of the separately obtained scales provides a test of the theory.

As has been seen, the logic of the present scaling technique consists in using the postulated behavior laws to induce a scaling on the dependent variable. In principle, this same approach could be applied to any theory which postulates quantitative laws of behavior although the detailed procedure would depend on the form of those laws and on the experimental design employed. The resulting scales would of course be correct only within the context of the postulated laws. However, they would have the considerable merit of being directly relevant to those postulates and of being directly associated with the development of the theory.

Finally, it should be noted that possession of a theoretically adequate scale of the dependent variable opens up the possibility of scaling the underlying variables. Miller's (1959) suggestion that a change in drive raises or lowers the whole gradient may be taken as one example. The difference between the scaled response before and after the change in drive should then represent the effect of that change in a theoretically adequate way. This would, for instance, give a basis for equating drives induced by different experimental operations.

POSSIBLE RELATIONS TO HULL-SPENCE THEORY

The purpose of this section is to explore one possible way in which the present formal model might be interpreted within a more general quantitative behavior theory. Such an interpretation is not necessary, of course, since the model is self-contained, but there are at least two reasons for so proceeding. An articulation of the model with an existing theory would presumably help to define the meaning of the model parameters. More importantly, such articulation would be expected to shed light on the validity of the assumptions and to indicate possible directions of change if change should prove necessary.

It is appropriate to restrict the discussion to general Hull-Spence theory. The treatment of this theory may seem somewhat casual but it should be realized that the intent is to indicate that an interpretation of the model within the theory may be possible and useful.

The following notation will be used. E, H, D, and K refer to reaction potential, habit strength, drive, and incentive motivation, respectively. Superscripts + and − index quantities associated with the approach and nonapproach responses, respectively. Thus the approach gradient is denoted by $E^+(x)$, and the avoidance gradient by $E^-(x)$.

Attention will be restricted to the following experimental situation. The subject is trained with a fixed reward and punishment until his behavior has stabilized. Theoretically, the values of $E^+(x)$, $E^-(x)$, $H^+(x)$, and $H^-(x)$ will then have reached asymptote. He is then given one of two instructions, either that the reward on the next trial will be greater (or less), or else that the punishment on the next trial will be greater (or less). Subscripts 1 and 2 will refer to quantities on the trial preceding and on the trial following this instruction, respectively.

Case a: Effect of Changed Reward

When the behavior has stabilized, the reaction potential difference corresponding to the difference between the gradients may, within Hull's theory, be written as:

$$E^+(x) - E^-(x) = K^+DH^+(x) - E^-(x). \quad [10]$$

The effect of the instruction will presumably be on K^+. Denote the new incentive motivation by αK^+ where α is constant. The reaction potential difference after the instruction is then:

$$E^+(x) - E^-(x) = \alpha K^+DH^+(x) - E^-(x). \quad [11]$$

To relate this to the model, set $f(x) = H^+(x)$, $a_1 = K^+D$, $a_2 = \alpha K^+D$, $b = c = 0$. The Hullian equations are thus seen to be consistent with the model, at least for this situation. Conversely, this application of Hullian theory is seen to justify A_1 as regards the approach gradient.

An analogous treatment in terms of Spence's (1956, p. 151) theory is more problematical. Apparently, K^+ should be considered as a function of x, but H^+ should be taken as independent of x. Although it would seem that D should be considered as dependent on x, it will be assumed here that D is the same at all distances from the goal, an assumption which will be given some justification below. These assumptions yield, analogous to Equation 10:

$$E^+(x) - E^-(x) = [K^+(x) + D]H^+ - E^-(x). \quad [12]$$

The effect of the instruction is again presumably to change the incentive motivation and it is assumed that the new incentive motivation is related to the old by the expression, $\alpha K^+(x) + \beta$, where α and β are constants. It will be noted that this assumption includes a number of special cases: additive, with $\alpha = 1$; multiplicative, with $\beta = 0$; and exponential growth with $0 \leq \alpha \leq 1$, $\beta = (1 - \alpha)K_M{}^+$, where $K_M{}^+$ is the maximum value of the incentive motivation. This yields, analogous to Equation 11:

$$E^+(x) - E^-(x) = [\alpha K^+(x) + \beta + D]H^+ - E^-(x) = \alpha H^+K^+(x) + \beta H^+ + DH^+ - E^-(x). \quad [13]$$

Setting $f(x) = K^+(x)$, $a_1 = H^+$, $a_2 = \alpha H^+$, $b_1 = 0$, $b_2 = \beta H^+$, and $c = DH^+$, it is seen that this interpretation of Spence's (1956) theory gives a result which is consistent with the model, at least for the given experimental situation. In particular, it justifies A_1 as far as the approach gradient is concerned.

Case b: Effect of Changed Punishment

In this case, the problem is whether an analysis of E^- can be made which is similar to the analysis of E^+ given in Case a. This does not seem possible within either Hull's or Spence's theory. For a threat of increased punishment would presumably increase the drive, D, and change the drive stimulus, S_D. The increased D would, as is well-known, theoretically act to increase the reaction potential difference, and hence also the speed of response, at all distances up to that point at which the subject no longer approaches. The changed drive stimulus would presumably have the opposite effect, of course, but unless the changes in D and S_D exactly balanced out in their effect on the approach tendency, the assumption of independence of the approach and avoidance gradients would be incorrect. It would thus appear that Miller's (1959) theory is, in this respect, inconsistent with the theories of Hull and of Spence.

There is, however, a possible alternative approach. This is to postulate a negative incentive motivation, K^-, which acts on H^- in the same way that K^+ acts on H^+, and to assume that the effect of the punishment instruction is on K^-. The theory would then be symmetrical with respect to approach and avoidance so that the analysis of E^+ in Case a would carry over directly to the analysis of E^- and exactly analogous conclusions would follow.

The concept of negative incentive motivation as used here would seem to merit some consideration. Such a concept has been used by Logan (1960, p. 222), and it is consistent with Miller's (1959) stimulus interpretation of drive. Indeed, if K^- were introduced, the concept of generalized drive, D, might be no longer necessary.

It should also be noted that, even if H^+ and H^- were constant at asymptote, K^+ and K^- might well still show trial-to-trial variation as a function of the sequence of reinforcing stimuli. If so, the learning formulation given in an earlier section would be directly applicable to the asymptotic data.

DISCUSSION

Any psychological theory has four aspects. First, a symbolic representation of the organism or, in less nominal terms, a delineation of the psychological processes underlying the behavior. Second, a set of laws relating these processes to their causal environmental determinants. Third, a set of combination rules which state how the underlying processes combine or coact to produce the observable behavior. Fourth, a set of measurement scales for the dependent and independent variables, and for the underlying processes. This fourth aspect is, of course, implicit in the first three but it will be useful to consider it in its own right.

The principal concern of the present model is with the last two aspects. The basic combination rule of the model is the assumption, A_2, that the overt response is proportional to the difference between the strengths of the two underlying processes, namely, the approach and avoidance gradients. By applying this assumption, under the general guidance of Assumption A_1, to various experimental situations, precise theoretical predictions are derived and serve as a test of the model. An important by-product, one which often results from a quantitative theoretical analysis, is the illumination of various experimental and statistical considerations of the kind discussed above which are involved in the detailed application of the model. This by-product would

seem the more valuable in that the relevance of a number of these considerations does not appear to be limited to the specific conflict model employed here.

It will be noted, however, that the exactness of prediction conferred by A_2 is not entirely satisfactory because of the arbitrary choice of a measuring scale for the dependent variable. Although response speed may be a reasonable approximation to a theoretically adequate scale, it would be optimistic to believe that it will be exact. Without a guarantee of A_2, however, it would be difficult to decide whether a deviation from prediction signals a basic flaw in the theoretical structure, or simply an inappropriate choice of scale of measurement. The difficulty is resolved by the scaling technique which yields a theoretically adequate scale of the dependent variable, i.e., a scale which is correct within the model employed.

The present development has little direct relevance to the first two aspects of theory listed above. The assumed underlying processes, namely, the approach and avoidance tendencies, are just those postulated by Miller (1959). Moreover, the scope of the model is narrow and formal in contrast to the substantive investigations of Miller and his colleagues which have been largely directed at the establishment of a set of causal laws, represented in particular by the assumption of gradient independence. Nevertheless, the model does bear indirectly on these two aspects. For, as has been seen, the success or failure of the model rests finally on Assumption A_1. If the model is rejected, then the assumption of gradient independence must be abandoned, or a different representation of the organism must be sought. In addition, it would seem that a theoretically adequate scale of the dependent variable is useful in defining the underlying processes and prerequisite to establishing scales for them.

Although much of the glamor of mathematical models seems to be associated with their predictive precision, this is not a good in itself. Prediction is usually only a means to the end of understanding behavior. Moreover, when there are several underlying processes, even a quantitative attack may yield only qualitative results. The advantage of a mathematical over a verbal formulation is that the former brings the basic ideas of the theory more quickly and more definitely to test. It is in this light that the present work should be viewed.

SUMMARY

This paper presents a mathematical model based on N. E. Miller's (1959) conflict theory and requiring minimal assumptions. An experimental situation is described and exact predictions are derived for it. A technique for obtaining a theoretically correct measurement scale for the dependent variable is also given. Finally, possible relations of the model to Hull-Spence theory are discussed.

REFERENCES

ANDERSON, N. H. An analysis of sequential dependencies. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory*. Stanford: Stanford Univer. Press, 1959. Pp. 248–264.

ANDERSON, N. H. Effect of first-order conditional probability in a two-choice learning situation. *J. exp. Psychol.*, 1960, **59**, 73–93.

ANDERSON, N. H. Two learning models for responses measured on a continuous scale. *Psychometrika*, 1961, **26**, 390–403.

BOWER, G. H. Choice-point behavior. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory*. Stanford: Stanford Univer. Press, 1959. Pp. 109–124.

BUSH, R. R., & MOSTELLER, F. *Stochastic models for learning*. New York: Wiley, 1955.

ESTES, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.

FISHER, R. A. The statistical utilization of multiple measurements. *Ann. Eugen., London*, 1938, **8**, 376–386.

HULL, C. L. *Principles of behavior*. New York: Appleton-Century, 1943.

HULL, C. L., FELSINGER, J. M., GLADSTONE, A. I., & YAMAGUCHI, H. G. A proposed quantification of habit strength. *Psychol. Rev.*, 1947, **54**, 237–254.

LEWIS, D. *Quantitative methods in psychology*. New York: McGraw-Hill, 1960.

LOGAN, F. A. Three estimates of differential excitatory tendency. *Psychol. Rev.*, 1952, **59**, 300–307.

LOGAN, F. A. *Incentive*. New Haven: Yale Univer. Press, 1960.

MARRIOTT, F. H. C. Tests of significance in canonical analysis. *Biometrika*, 1956, **39**, 58–64.

MILLER, N. E. Liberalization of basic S-R concepts. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. 2. New York: McGraw-Hill, 1959. Pp. 196–292.

MUELLER, C. G. Theoretical relationships among some measures of conditioning. *Proc. Nat. Acad. Sci.*, 1950, **36**, 123–130.

RAO, C. R. *Advanced statistical methods in biometric research*. New York: Wiley, 1952.

SEWARD, J. P. The constancy of the I-V: A critique of intervening variables. *Psychol. Rev.*, 1955, **62**, 155–168.

SPENCE, K. W. The relationship of response latency and speed to the intervening variables and N in S-R theory. *Psychol. Rev.*, 1954, **61**, 209–216.

SPENCE, K. W. *Behavior theory and conditioning*. New Haven: Yale Univer. Press, 1956.

SPENCE, K. W. *Behavior theory and learning*. Englewood Cliffs: Prentice-Hall, 1960.

(Received January 9, 1961)

Psychological Review
1962, Vol. 69, No. 5, 415–427

FROM ASSOCIATION TO STRUCTURE[1]

GEORGE MANDLER

University of Toronto

The dialogue between stimulus-response theories and cognitive theories has antecedents going back at least as far as the beginnings of scientific psychology. Protestations that the conflict is imaginary or merely semantic are unlikely to eliminate fundamental theoretical differences. To argue that we are dealing with the same data regardless of theoretical predilections (e.g., Kendler, 1952) would sound strangely incongruous in the context of other sciences. The question "What is the center of the universe?" was not a theoretical blind alley, and neither is the question "What is learned?" as Campbell (1954) and Rozeboom (1958) have shown rather convincingly. In the present paper we shall examine briefly the major differences between the two positions with regard to the definition of response or output and then proceed to illustrate a possible locus of transition from associative to cognitive or structural processes.

The distinction between associationist and cognitive views may be described as a difference between a topographic and a categorical response definition. While the associations of the associationistic schools are usually considered to be between discrete events in the environment (the stimuli) and topographically well delimited and discrete events associated with the organism (the responses), the cognitive theorist is more likely to talk about formal categories of behavior, giving relatively less attention to the topographic aspects of the behavior classed within those categories. The important questions are: Do organisms learn generalizable, but discrete, responses in specific situations or are rules of behavior, maps, or schemata laid down which connect various behaviors and environmental inputs? Do organisms learn what to "do," or do they learn "what leads to what"? While the distance between the two views has been reduced somewhat by suggestions such as Logan's (1960), who aggregates responses as a function of differential reinforcement, the problem of response definition still appears to be basic.

The so-called cognitive theories belong to a much larger class of theories which have claimed to be "structural," including not only the gestalt school and its heirs but also the speculations of Piaget, Bartlett, and Hebb, among others. What these positions have in common is a postulation of the organization of behavior not derivable from any combination or association of stimulus-response links.

Rather than reserving the construct of "structure" for this kind of theory, we would prefer a more general statement of the structural position. *Structures are temporal and probabilistic linkages of inputs and behavior which are available in functional units.* These

[1] This paper was prepared in connection with Research Grant M-4852 from the National Institute of Mental Health, United States Public Health Service. The section on overlearning and transfer was presented at the American Psychological Association Symposium on Transfer of Training in September 1957. The initial version of the paper was prepared while the author was a Fellow at the Center for Advanced Study in the Behavioral Sciences, Stanford, California. He is indebted to William Kessen, who contributed not only valuable critical goads but also additions of substance, and to Jean Mandler for her many suggestions.

units—broadly conceived—may be as simple as a reflex arc, as associationistic as a habit-family hierarchy, or as cognitive as a means-end relationship. The output of a structure may be topographically limited or categorically defined; the question to be addressed concerns the conditions under which structures are developed and the relation between structures which mediate topographically limited behavior and those which mediate behavior described in categorical form.

The major notion to be presented in this paper and applied to a limited set of phenomena is the proposition that an associationist and a cognitive view of behavior may work side by side—that they are compatible. More specifically, it will be maintained here that there is now enough evidence available to suggest that, in some limited cases, cognitive characteristics of the organism may be developed out of associationist processes.

FROM ASSOCIATION TO COGNITION

The broad area of evidence to which we will appeal in support of such a proposition covers the warm-up effect and the development of learning sets. More specific evidence will be developed out of a series of studies, strewn through the literature, on the effects of overlearning on transfer of training.

We can briefly make some distinctions between three types of effects:

The warm-up effect. Prior experience in the particular setting of an experiment results in an increasing facilitation on the criterion task as a function of amount of such experience (Mandler, 1956; Thune, 1950). Thus Irion (1948) described the warm-up effect in general; while Thune (1950) demonstrated that both color-guessing and paired-associate learning were equally effective in facilitating learning of a test list. Hamilton (1950) suggested that these effects may be relatively transitory in nature.

Learning to learn. When a subject is given repeated and related tasks, there is an increasing facility in the performance of these tasks. While this effect has been most dramatically demonstrated by Harlow (1949), Bair (1902) in a classic study was able to conclude that "learning to do one thing gives us the capacity for learning to do other things" (p. 45). Most recently, Duncan (1960) showed that performance improved markedly as a function of number of prior tasks.

Overlearning and transfer. High degrees of learning of a particular response lead to positive transfer to an incompatible response, even though lower levels of learning produce negative transfer (cf. Reid, 1953).

These three general effects all have one characteristic in common, *practice on one task may have a facilitating effect on another task even though the behavior required in the two tasks is topographically different, and may even be incompatible.*

To what extent can current associationist stimulus-response theories accommodate to these phenomena? As far as the warm-up effect is concerned, it seems to be reasonable to assume that within broad limits a warm-up task permits the subject to "assume more rapidly the postural and 'attentive' adjustments necessary for optimal performance" (Thune, 1950, p. 438). However, when we deal with learning sets and overlearning, a topographically limited stimulus-response theory sometimes fails to do justice to the phenomena.

The general position to be adopted here in respect to learning sets and overlearning is similar to that espoused originally by Harlow (1949) and ex-

panded subsequently (1959).[2] Harlow suggested that the formation of hypotheses and insight is a function of the continuous development of learning sets, rather than the discontinuous appearance and replacement of cognitive structures. The writer has previously suggested (Mandler, 1954a) that during overlearning, i.e., continuous errorless performance, there are developed symbolic analogues—we now prefer to call them analogic structures—of the overt responses. The process unfolds in the following manner: First, the organism makes a series of discrete responses, often interrupted by incorrect ones. However, once errors are dropped out and the sequence of behavior becomes relatively stable—as in running a maze, speaking a word, reproducing a visual pattern—the various components of the total behavior required in the situation are "integrated." Integration refers to the fact that previously discrete parts of a sequence come to behave functionally as a unit; the whole sequence is elicited as a unit and behaves as a single component response has in the past; any part of it elicits the whole sequence. This process of integration has been called "response learning" (Mandler, 1954a) and may be applied to any consistent responding of whatever topography and complexity. Thus, pressing a button is integrated as much as running a maze or speaking a sentence. Once a response sequence has been integrated and acts as a unit, it develops a structural representation, a "central" analogue of this new response unit which can function independently of the overt response sequence.

The notion of integration deals in part with the problem of defining units of behavior. When complex behavior is built up from previously disparate units, much trial and error behavior may characterize the slow development of the new, larger unit. The fact that the components may themselves have been overlearned—and structurally represented—lends a peculiarly "cognitive" flavor to the explicit or implicit hunt and peck search for the appropriate components. However, once the units have been selected and the new sequence has been integrated, it tends to be elicited as a whole and has a quasi-automatic appearance in the way it is evoked by the appropriate situational cues. It is this development from unsure steps to the well integrated sequence that characterizes learning to drive a car, to type a word, or to run a maze. Such a process is related to Miller's (1956) discussion of the chunking of behavior and probably led to Tolman's (1933) conclusion that "the Sign-gestalt formula is appropriate only for the early stages of learning and not for the later ones" (p. 254).

Analogic structures permit covert trial and error behavior, i.e., cognitive manipulation of previously established behavior. In this sense, the analogic representation of a prior behavior sequence is one possible "hypothesis" to be applied to a particular situation. Given many such structures which are relevant to a situational input, the several structures will occur seriatim and covertly until an appropriate one is expressed behaviorally. If it is incorrect, a new structure may occur, thus giving the appearance of an entirely discontinuous process, of shifting from hypothesis to hypothesis. We say "appearance" of discontinuity because a continuous process prior to the appearance of cognitive structures in fact gives rise to them. Structures are developed on the basis of associationist stimulus-re-

[2] The point of view presented here has familial ties with Hebb's (1949) theory, with some of Piaget's (1952) ideas, and with a recent essay by Miller, Galanter, and Pribram (1960).

sponse relationships but, once established, enable the organism to behave "cognitively." In Harlow's (1959) words: ". . . it is the amount of intraproblem learning which determines the interproblem transfer" (p. 502).

In the learning set studies, the problem is not one of covert trial and error, i.e., the choice between two incompatible or antagonistic responses, but rather one of positive transfer from one problem to the other when the tasks have formal similarities, but no overlap in content. What will be suggested here—and this is one of the reasons why we should move from specific symbolic analogues to more general structures—is that in any one task, e.g., in discrimination learning, structures (or learning sets) are developed which are formally identical with those of the succeeding and preceding tasks. Thus, for example, the organism in choosing the "odd" member of a set of three objects engages at first in overt behavior which has the following characteristic: Take three objects, A, B, and C; the organism physically examines all three objects in some order, say A, B, C (i.e., he makes differentiating responses to them—Mandler, 1955—as he actually manipulates them); whenever the differentiating response to A is similar to the differentiating response to B (they are "alike") then a consummatory response will follow the choice of C (C is the odd object and its choice is reinforced); if the differentiating response to B is different from that made to A, then B will be followed by a consummatory response; if the differentiating response to A is followed by a consummatory response, no further behavior is elicited. In all cases the consummatory response is the stop rule. At this point we are still dealing with overt behavior during the early stages of learning. However, as the sequence of differentiating and overt responses is integrated, a structure of that sequence will be developed. This structure has three nodal points each of which represents a differentiating response and a choice response, none of which needs to be made overtly and the sequence can now run off in the following manner: Look at one object and make a differentiating response, look at the second object and, if the same differentiating response occurs as the one made to the first, then make an overt choice of the third object. Note that the matching requirement need not be held for more than two objects. Assume that A, B, and C have the characteristics k, k, and l, respectively. C therefore is the odd object. If the subject first examines A and then B, then the object following a match is the correct one; if he examines them in the order A, C, and B then a repetition of the sequence will lead to the same occurrence of a choice following a match (he has examined A, k; C, l; B, k; A, k; and now chooses C); if the order is C, A, and B, a return to C will lead to a similar result. In other words, the only required structures are an orderly sequence of covert choices and an overt choice or act following two similar differentiating responses. What the analogic structure permits the organism to do is to eliminate the overt examination of *all* the objects in order to arrive at the consummatory response.[3]

This exercise has been used to demonstrate *one way* in which cognitive categorical behavior may be slowly built up out of regular overt behavior sequences. One of the major variables of associative theories—frequency—appears to be an important antecedent of simple structures.

The development of analogic struc-

[3] The similarity between the process described here and the simulation of cognitive processes with computers (e.g., Newell & Simon, 1959) is obvious.

tures, and the resulting cognitive behavior is particularly relevant to verbal behavior which, in the natural setting, demonstrates high degrees of learning as well as cognitive correlates. And finally, while the apparent cognitive sequelae of overlearning have been demonstrated in mammals (monkeys, men, and rats), studies by Warren and his associates (Warren, 1960; Warren, Brookshire, Ball, & Reynolds, 1960) fail to find these effects in fish and chicks. More comparative data will have to be collected before we can speculate about the phylogenesis of cognition.

OVERLEARNING AND TRANSFER

In applying the notions developed above to the effects of overlearning on transfer of training, we shall proceed cautiously before invoking cognitive or structural explanations. We shall first review some of the typical findings in the literature, then outline the major variables apparently involved in the phenomena, and finally examine the extent to which associative factors may account for them, before invoking analogic structures as operative in these situations. The major independent variable with which we will be concerned is the degree of training on the original task, in particular when such training goes beyond mere mastery of the original task. We will be primarily interested in transfer situations which exemplify the Müller-Schumann paradigm, the A–B, A–C situation where identical or highly similar stimuli are used in both tasks but where the required response varies from the original to the transfer task. Overlearning beyond the attainment of asymptotic behavior provides repeated trials which are topographically highly similar from occasion to occasion, i.e., the same sequence of behavior occurs again and again. However, in some situations such as the learning of a complex maze by a rat or the learning of a paired-associate list by a human subject, it is difficult to specify when asymptote has been reached. In the case of the rat, some choice points will be learned very early and while the animal may still be making errors at some other choice points, he is at the same time overlearning the earlier acquisitions. In the case of verbal learning, e.g., of a list of 10 paired associates, the subject may be overlearning some pairs by the time he reaches even a low criterion of 3 or 4 correct out of 10. In this case, learning the list to a criterion of no errors will include some overlearning of some of the pairs in the list. Therefore, some data from the literature will be presented which are concerned with relatively lower degrees of training of the entire task but which probably include overlearning of some parts.

First, we can examine some typical results from the literature on the effects of degree of A–B training on A–C performance. The degree to which any one of these examples conforms to the ideal A–B, A–C situations does, of course, vary. Such variation extends from the ideal, where A is in fact an identical stimulus situation in both original and transfer tasks, to some degree of similarity between the original and transfer task. Similarity may imply the presence of common elements in the two situations or other constants across situations which are likely to elicit the original response.

Figure 1 shows some typical results from both animal and human studies. In each case the learning of the transfer situation with no prior training has been used to estimate the point of zero transfer and amount of transfer is plotted against number of trials on the original task. Performance for groups that have had some original learning

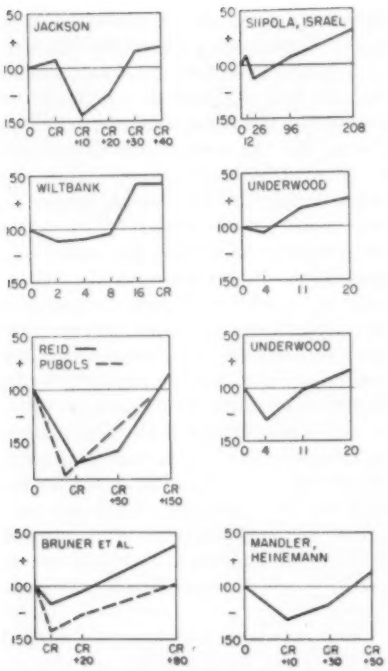

FIG. 1. Transfer of training as a function of degree of original learning. (See text for explanation.)

prior to the transfer task is represented in terms of percentages of the performance of the group with no prior training.

One of the most striking aspects of these graphs is the similarity of the general function of the relationship between transfer and degree of prior training. An initial negative transfer effect is followed by either a return to zero transfer, or the appearance of large positive transfer effects. The effect is **U** shaped in all instances.

The eight graphs in Figure 1 represent animal studies on the left and human studies on the right. Jackson (1932) trained his first group of rats to a criterion of 4 out of 5 correct trials on a multiple-cul maze. His second, third, fourth, and fifth group learned to criterion plus 10, 20, 30, and 40 overlearning trials, respectively. Transfer to a similar maze shows an initially large negative transfer effect followed by a slight positive effect with high degrees of overlearning. Wiltbank (1919), in a study using similar apparatus but lower degrees of learning (0, 2, 4, 8, 16 trials, and a mastery criterion), obtained highly similar results, though less negative and more positive transfer.

The next graph represents Reid's (1953) study, in which animals learned a black-white discrimination in a **Y** maze. They show negative transfer when reversed to the previously nonreinforced stimulus *if* they were trained only to criterion on the first task. However, with 50 overlearning trials we find less negative transfer, and positive transfer for the group that had 150 overlearning trials. Pubols (1956), in replicating this particular study, finds a similar effect though no positive transfer for the highly overlearned group. Galanter and Bush (1959) have presented some data on reversal learning which fail to show any positive transfer following high degrees (144 trials) of overlearning. Their animals, however, were run under conditions of distribution which resulted in much more spaced trials than Pubols used. Such a condition might be unfavorable to the development of stable structures.

The last graph for the animal studies represents a study by Bruner, Mandler, O'Dowd, and Wallach (1958). In this study animals were trained to learn a left-right alternation sequence in a four-unit maze. After varying degrees of acquisition in this original task, they were switched to the mirror image of the original task, e.g., the right-left-right-left response. Animals either learned to a criterion of 80% correct or were given 20 and 80 overlearning

trials, respectively. Once again we find an initial negative transfer effect followed by a large positive transfer effect with 80 overlearning trials. The second (dotted) curve in this graph represents a group which was run when 36 hours hungry, the first group was run when 12 hours hungry.

We can now turn to the human studies. The first example is a study by Siipola and Israel (1933) in which subjects learned various telegraph codes to letters of the alphabet as stimuli. This study is not strictly an A–B, A–C situation. It involved the recombination of stimuli and responses in the transfer task. The same stimuli and same responses as in the original task were used in the transfer task with different combinations of stimuli and responses. Four groups were given different degrees of training; their means were 12, 26, 96, and 208 trials on the original task, respectively. The last two groups received high degrees of overlearning. In this particular case there is a slight positive transfer effect for the first group which is followed by a negative transfer effect and then by high positive effects for the last two groups.

Underwood (1949) used lists of 10 paired-associate adjectives. The most marked effect is the high degree of positive transfer with high degrees of overlearning, with only little negative transfer for the group receiving the least amount of training on the original task. In this particular situation, the first group had 3 correct out of 10 pairs, the second 8 correct out of 10 pairs, and the third group one complete correct trial plus five additional trials. The absence in this particular case of any large negative effect may be due to the lack of a warm-up control which we will discuss presently. Underwood also gives the number of errors that were made during the first two anticipation trials (Trials 2 and 3) of the transfer task. These are shown in the next graph where we find negative transfer effects for the least trained group which change to zero or positive transfer for the more highly trained group.

The next graph shows a study by Mandler and Heinemann (1956) in which subjects learned three-consonant nonsense syllables as responses to one digit numbers. The first overlearning group was given 10 trials beyond complete mastery of the list, the second 30 overlearning trials, and the third 50 overlearning trials. Again we find initial negative transfer effects followed by zero or positive transfer.

In examining these findings, five general processes will be considered: warm-up and learning set effects, response competition, stimulus discrimination, response learning, discontinuous structural factors.

Warm-up and learning set effects have been grouped together because—in the kinds of design presented here—they are usually confounded. Thus, it is often not determinable to what extent an animal's prior experience in a maze produces varying degrees of specific postural and attentive habits (warm-up) as against nonspecific structural effects (learning set). Presumably both kinds of effects are operating and may, for present purposes only, be considered together. But in order to evaluate the effects of overlearning on transfer, these more general experiential factors should be separated somehow from factors assignable to the competition and facilitation between original and transfer responses. We can examine the ideal model and determine what kind of control groups are necessary to separate situational and nonspecific transfer effects from those transfer effects which are peculiar to the juxtaposition of the

particular original and transfer tasks. The model can be conceptualized as follows:

	Original task	Transfer task
Ideal paradigm	A–B	A–C
Control condition	D–E	A–C

The first line, A–B, A–C, is the ideal paradigm. However, any experience with the general class of tasks which A–C represents may result in transfer effects not due to A–B alone. In the second line the necessary control conditions are shown. In this case a D–E situation is necessary, which must be learned to the same degree as A–B was learned but which contains neither the stimuli of the A–C situation nor the responses of either the A–B or the A–C situation. In simple paired-associate learning this is not difficult to accomplish. Obviously we can have control groups which learn a series of paired-associate adjectives as in A–B, but with adjectives in the stimulus and response position which are different from either A, B, or C. A brief example of what happens when one does or does not use controls for this kind of situation is shown in Figure 2 which recapitulates the data from Mandler and Heinemann (1956) but incorporates a control for warm-up and learning-set effects. The first graph of the figure shows the raw data for the A–B, A–C situation, and in the second graph is shown the raw data for a D–E, A–C situation. (For comparison purposes the large positive effects of an A–B, C–B condition are also shown.) In both cases we find increasing positive transfer as a function of overlearning on the original task. However, the positive transfer for the control condition is greater for the first two groups and less so for the last group. The resultant is shown in the third graph, which shows initial negative transfer followed by positive transfer.

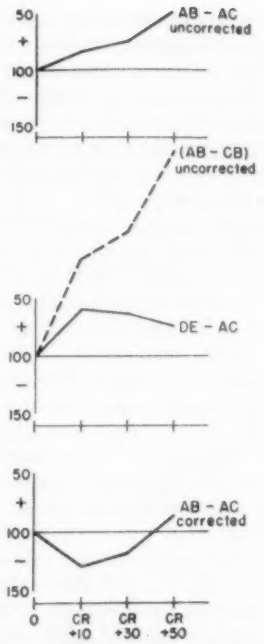

FIG. 2. Effect of warm-up correction on transfer effects. (See text for explanation.)

The paired-associate situation is relatively easy to consider in controlling for warm-up and learning to learn effects. It becomes more difficult when we think about some of the animal studies. It may be difficult to specify a warm-up condition which will separate practice on a specific maze from general maze or situational experience, but such control groups seem to be essential.

Some of the studies do suggest possible control groups. In Reid's (1953) study, for example, it might be possible to use control groups who learn to discriminate but do not learn a differential black-white response. Reid argues for the acquisition of a new response, the response of discriminating, which does not occur until late in overlearning. If animals were to learn a discrimination

between a triangular and a circular stimulus and given various degrees of training on this discrimination before being transferred to a black-white discrimination, we would then have a control group which does not have the black-white practice, but does have practice in discriminating. Any resultant advantage which still accrues to the group that had its prior practice on the black-white discrimination would be specific to that particular situation and not related merely to being in a discrimination task. In general there tends to be some confusion between specific transfer effects and general transfer effects due to warm-up or learning to learn. Transfer studies must find some way of separating general from more specific transfer effects.

The first of the specific transfer effects to be discussed is the classical one of response competition or interference. The response which was appropriate in the original task is elicited in the transfer task, particularly if the stimulus situation in the two tasks is highly similar. This effect has led past investigators to call the A–B, A–C paradigm the negative transfer paradigm par excellence. Despite the failure to find such a result in the overlearning studies, the available evidence in transfer studies suggests that the interference or competition hypothesis need not be abandoned. What appears to be the case is that interference effects occur very early during the transfer task and that "factors making for positive transfer . . . do not gain the ascendancy over the negative tendencies until the latter stage of the learning process" (Wiltbank, 1919, p. 42).

The important point is that the probability that Stimulus A, which is present in both situations, will elicit Response B, which was learned in the original task and which interferes with the performance of Response C required in the transfer task, does increase as a function of prior A–B pairings. But, despite this increase in intrusion errors, the overall efficiency in learning the transfer task is increased.

Underwood's (1949) data show large interference effects during the first two anticipation trials, probably primarily due to intrusions of the now erroneous response during early learning of the transfer task. In animal studies, Reid (1953) has shown that his highly overtrained animals make many more intrusion errors, i.e., going to the now incorrect stimulus, than do the less trained animals. The Bruner et al. (1958) study also indicates that the amplitude (in time) of the now incorrect "left" response at the first choice point of the transfer task increases as a function of degree of original training—though in the 12-hour deprived group only.

As far as stimulus differentiation is concerned, the more experience the subject has had in differentiating the stimuli used in transfer, the less likely we are to find intratask competition among various stimuli (Gibson, 1940, 1953). But whenever the stimulus situation in the two tasks is identical, as in the A–B, A–C paradigm, stimulus differentiation is only relevant to the degree that the subject has learned a consistent response to the stimulus in the original task—nothing new needs to be learned about the stimulus in the transfer task.

More important than stimulus learning is response learning. The degree to which the organism "knows the response" is an important factor in his performance of the transfer task. If that task requires a new response which has not been previously practiced and integrated, the early trials of the transfer task will be devoted to response learning as well as to the elimi-

nation of competition with the responses of the original task.[4] Special attention must be paid here to extra-experimental experience. Responses may well have been learned prior to the transfer task in situations other than the original task. For example, in paired-associate learning with adjectives it is obvious that whatever adjectives are used in the transfer task, the subject has had some prior experience in learning how to make the responses—a condition not applicable to some nonsense syllable learning experiments where the subject has to learn the response *de novo*.

Before considering discontinuous cognitive effects, we shall summarize the hypothesized effects of the other four factors on the transfer task as a function of degree of training on the original task.

1. We may tentatively assume that warm-up activity and learning sets both have a monotonic facilitating effect on subsequent behavior.

2. Degree of prior learning of an "incompatible response" has a monotonically increasing interfering effect. It also appears to be the case that these effects, while manifested in intrusion errors, do not necessarily hinder the rapid acquisition of the new response.

3. Stimulus discrimination also seems to have a monotonic facilitating effect on subsequent learning.

4. Finally, response learning facilitates subsequent acquisition of a new response to the extent to which the required response has been previously practised.

There seems to be some evidence that Bair's (1902) and Wiltbank's (1919) conclusion that negative effects (response competition in our terms) are maximal during early learning and that positive effects do not appear until the later stages of learning is essentially correct. What may also be the case is that during the later stages of acquisition or performance new responses are learned which can be transferred to the new task. The "observing response" or "response of discriminating" described by Wyckoff (1952), Reid (1953), Pubols (1956), and others, represents the learning of a new response which is easily transferred to the new situation and which, as Pubols suggests, is acquired during the later stages of the learning process. However, a recent study by Birch, Ison, and Sperling (1960) using single stimulus presentations showed essentially similar results for the effects of overlearning on discrimination reversal. They conclude that their findings "effectively rule out interpretations . . . in terms of responses of discriminating simultaneously presented discriminanda" (p. 40). Birch et al. also appear to dispose of the suggestion by D'Amato and Jagoda (1960) that "overlearning should not facilitate reversal learning in a successive discrimination situation, since avoidance of S_- is maintained by the very nature of the situation" (p. 259).

If we apply these various factors to the studies summarized in Figure 1, one possible conclusion is apparent. If negative factors such as response competition are dominant after low degrees of original learning, and if positive factors such as warm-up and learning set do not appear in full force until later stages of learning, i.e., during overlearning, then an initial negative transfer effect followed by a reversal in a positive direction, after the organism has had longer experience with the original task, is predictable. Such an explanation need not invoke any cognitive discontinuous processes. Wilt-

[4] Underwood and Schulz (1960) have elaborated earlier suggestions concerning the importance of this variable in verbal learning.

bank's, Jackson's, Reid's, and Pubols' studies all might be explained in terms of specific learning occurring during the original task. The experiment by Bruner et al. (1958) designed specifically to demonstrate the acquisition and use of rules, is probably also a case of response transfer, since an integrated right-left-right-left sequence could easily be transferred to a left-right-left-right sequence—large chunks of the required behavior are practiced during original learning and probably well integrated during overlearning of the original task. A more crucial question would be raised if a control group similar to the one suggested for the Reid study were included. If animals were trained on a circle-triangle discrimination and then transferred to a black-white discrimination,, their behavior on the black-white task would provide a baseline for evaluating reversal effects. Thus, once reversal data are adjusted for the effects of varying degrees of prior maze experience, it will be possible to state whether the overlearning reversal effects can be explained in terms of maze-specific behavior or whether animals during overlearning acquire something other than warm-up, learning sets or discriminating responses, namely some cognitive structure about black and white stimuli.

In the situation where the subject learns to make a new response to an old stimulus, the mere learning of the old response is irrelevant. If we further assume that stimulus discrimination is not a major factor contributing to transfer, we are left with three effects. Response competition, the major negative effect, should increase to asymptotic levels as a function of prior pairings of old stimulus and old response. Warm-up and learning-set effects should increase more markedly during the later (overlearning) trials and produce positive transfer effects. The learning of new nonspecific responses (e.g., observing responses) probably also appears late during original learning and has positive effects. However, once the appropriate control groups are used, all of the positive effects should be controlled and we should be left only with the negative effect of response competition. When these negative effects change to zero or positive effects despite the use of the appropriate controls (as in Mandler, 1954b; Mandler & Heinemann, 1956; and Siipola & Israel, 1933; and possibly in Underwood, 1949) it is appropriate to invoke some other hypothesis. We have suggested that overlearning experience with the old response and the formation of an analogic structure permits the subject to manipulate that response, to "think about" the problem without making overt errors. Just as we assume that the animal in the oddity problem can use cognitive structures to eliminate the need for overt choices, and in discrimination problems can manipulate black-white choices covertly by inhibiting responses to the "incorrect" stimulus, the subject in the A–B, A–C situation may be able to eliminate overt B responses.

Starting off with the assumption of a continuity from simple associative structures to cognitive structures, we have examined one area of research in detail to find evidence for the appearance of cognitive structures. Our investigation has uncovered a variety of associative factors which are adequate to explain what may at first glance appear to be discontinuous behavior. This will, of course, be no surprise to the committed S–R theorists who have always maintained that most, if not all, observations on apparently cognitive, categorical behavior will, in the long run, be amenable to an associationistic explanation. While we would agree that sophisticated associationist posi-

tions (e.g., Miller, 1959; Spence, 1956) have been too readily abandoned by their critics, we would maintain that some behavioral phenomena defy current associationistic interpretation. However, recourse to cognitive structures should await the verdict of the laboratory; the proper use of some of the control conditions suggested above and a more fine-grained analysis of warm-up effects and learning sets will determine which of the phenomena discussed here will ultimately call for cognitive explanations.

SUMMARY

Given the distinction that associationist and cognitive theories make in respect to the definition of response, we have examined the proposition that one source of cognitive behavior may be found in associative antecedents. The argument for a transition from associations to cognitions was buttressed by reference to the development of learning sets and developed out of the notion of response integration and the emergence of analogic structures. Empirical evidence on the effect of overlearning on transfer of training was examined in detail to differentiate findings that seem to be amenable to associative explanations from those that seem to call for structural, cognitive explanations.

REFERENCES

BAIR, J. H. The practice curve: A study in the formation of habits. *Psychol. Rev. Monogr. Suppl.*, 1902, **5** (2, Whole No. 19).

BIRCH, D., ISON, J. R., & SPERLING, SALLY E. Reversal learning under single stimulus presentation. *J. exp. Psychol.*, 1960, **60**, 36–40.

BRUNER, J. S., MANDLER, JEAN M., O'DOWD, D., & WALLACH, M. A. The role of overlearning and drive level in reversal learning. *J. comp. physiol. Psychol.*, 1958, **51**, 607–613.

CAMPBELL, D. T. Operational delineation of "what is learned" via the transposition experiment. *Psychol. Rev.*, 1954, **61**, 167–174.

D'AMATO, M. R., & JAGODA, H. Effects of extinction trials on discrimination reversal. *J. exp. Psychol.*, 1960, **59**, 254–259.

DUNCAN, C. P. Description of learning to learn in human subjects. *Amer. J. Psychol.*, 1960, **73**, 108–114.

GALANTER, E., & BUSH, R. R. Some T-maze experiments. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory*. Stanford: Stanford Univer. Press, 1959. Pp. 265–289.

GIBSON, ELEANOR J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.

GIBSON, ELEANOR J. Improvement in perceptual judgments as a function of controlled practice and training. *Psychol. Bull.*, 1953, **50**, 401–431.

HAMILTON, C. E. The relationship between length of interval separating two learning tasks and performance on the second task. *J. exp. Psychol.*, 1950, **40**, 613–621.

HARLOW, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, **56**, 51–65.

HARLOW, H. F. Learning set and error factor theory. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. 2. New York: McGraw-Hill, 1959. Pp. 492–537.

HEBB, D. O. *The organization of behavior*. New York: Wiley, 1949.

IRION, A. L. The relation of "set" to retention. *Psychol. Rev.*, 1948, **55**, 336–341.

JACKSON, T. A. General factors in transfer of training in the white rat. *Genet. psychol. Monogr.*, 1932, **11**, 1–57.

KENDLER, H. H. "What is learned?" A theoretical blind alley. *Psychol. Rev.*, 1952, **59**, 269–277.

LOGAN, F. A. *Incentive*. New Haven: Yale Univer. Press, 1960.

MANDLER, G. Response factors in human learning. *Psychol. Rev.*, 1954, **61**, 235–244. (a)

MANDLER, G. Transfer of training as a function of degree of response overlearning. *J. exp. Psychol.*, 1954, **47**, 411–417. (b)

MANDLER, G. Perceptual learning: Stimulus differentiation or differentiating response? *Psychol. Rep.*, 1955, **1**, 79–82.

MANDLER, G. The warm-up effect: Some further evidence on temporal and task factors. *J. gen. Psychol.*, 1956, **55**, 3–8.

MANDLER, G., & HEINEMANN, SHIRLEY H. Effect of overlearning of a verbal response on transfer of training. *J. exp. Psychol.*, 1956, **51**, 39–46.

MILLER, G. A. The magical number seven, plus or minus two: Some limits on our

capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97.

MILLER, G. A., GALANTER, E., & PRIBRAM, K. H. *Plans and the structure of behavior.* New York: Holt, 1960.

MILLER, N. E. Liberalization of basic S–R concepts: Extensions to conflict behavior, motivation and social learning. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959. Pp. 196–292.

NEWELL, A., & SIMON, H. A. *The simulation of human thought.* Santa Monica, Calif.: Rand, 1959.

PIAGET, J. *The origins of intelligence in children.* New York: International Univer. Press, 1952.

PUBOLS, B. H. The facilitation of visual and spatial discrimination reversal by overlearning. *J. comp. physiol. Psychol.*, 1956, **49**, 243–248.

REID, L. S. The development of noncontinuity behavior through continuity learning. *J. exp. Psychol.*, 1953, **46**, 107–112.

ROZEBOOM, W. W. "What is learned?" An empirical enigma. *Psychol. Rev.*, 1958, **65**, 22–33.

SIIPOLA, ELSA M., & ISRAEL, H. E. Habit interference as dependent upon stage of training. *Amer. J. Psychol.*, 1933, **45**, 205–227.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

THUNE, L. E. The effect of different types of preliminary activities on subsequent learning of paired-associate material. *J. exp. Psychol.*, 1950, **40**, 423–438.

TOLMAN, E. C. Sign-gestalt or conditioned reflex. *Psychol. Rev.*, 1933, **40**, 246–255.

UNDERWOOD, B. J. Proactive inhibition as a function of time and degree of prior learning. *J. exp. Psychol.*, 1949, **39**, 24–34.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

WARREN, J. M. Reversal learning by paradise fish (*macropodus opercularis*). *J. comp. physiol. Psychol.*, 1960, **53**, 376–378.

WARREN, J. M., BROOKSHIRE, K. H., BALL, G. G., & REYNOLDS, D. V. Reversal learning by white leghorn chicks. *J. comp. physiol. Psychol.*, 1960, **53**, 371–375.

WILTBANK, R. T. Transfer of training in white rats upon various series of mazes. *Behav. Monogr.*, 1919, **4**, No. 17.

WYCKOFF, L. B. The role of observing responses in discrimination learning. *Psychol. Rev.*, 1952, **59**, 431–442.

(Received February 15, 1961)

Psychological Review
1962, Vol. 69, No. 5, 428–449

STIMULUS–INTENSITY EFFECTS IN RESPONSE EVOCATION[1]

R. A. CHAMPION[2]

University of Sydney

It is a well-established fact that the strength of a response varies directly with the intensity of its eliciting stimulus. The relationship holds for a wide variety of reactions and within the range of stimulus intensities from the lower absolute threshold to the level at which the receptors suffer physical damage. Despite the widespread recognition of the phenomenon, two basic problems remain unsolved. First, it is not known exactly what property of the stimulus situation causes the effect, i.e., there is no precisely articulated theory of stimulus-intensity effects which is firmly established. Second, the question arises as to whether the amount of true *learning*, as distinct from performance, varies with the intensity of the eliciting stimulus during training. Beginning with an outline of the two theories advanced to explain stimulus-intensity effects, the following discussion deals briefly in that context with attempts to assess the role of stimulus intensity in learning; the greater part of the paper is then taken up with a consideration of some empirical evidence on response evocation which seems to allow a choice between the two theories.

[1] Part of the apparatus used in these studies was provided by Commonwealth of Australia Research Grant No. 2002.

[2] The author is indebted to P. G. Ward for assistance in interpreting the data of Experiment II and to Joan E. Jones for assistance in gathering and treating the data of Experiments III and IV.

Theories of Stimulus-Intensity Effects

The first theorist to deal with the stimulus-intensity problem in formal terms was Hull (1951). In typical fashion he began by making it the subject of a postulate and, in effect, assumed a constant relationship between response strength and stimulus strength, other things being equal. He named the effect "stimulus-intensity dynamism" and represented it in his behavior system with the intervening variable V. "Dynamism" then became the stimulus-intensity component of reaction potential, acting multiplicatively with the other positive factors in the system. Hull (1952) later proceeded beyond the mere assumption of a relationship to a more elaborate theory in which he attempted to explain the effect with the concept of the molar stimulus trace. According to this formulation, the physiological effect of stimulation, occurring when a brief stimulus impinges upon a suitable receptor, reaches a maximum about 450 milliseconds after the beginning of the stimulus. Hull thereby limited the stimulus-dynamism process to the *onset* of a stimulus and to short-latency responses such as the eyeblink. He also gave the appearance of allowing for the effect of stimulus intensity on amount learned by distinguishing between V_1, the dynamism of the stimulus during the *learning* of the response, and V_2, the dynamism of

the stimulus during the later *evocation* of the response. At one point Hull (1951) did suppose that *habit strength* during learning depended on V_1, but he later (1952) shifted to the idea that *reaction potential* during training involved V_1 as a component. It would appear that, in his later theorizing, Hull confined V to the strength of the signaling stimulus trace as dependent on the age of the trace instead of on the intensity of the stimulus. Thus he implied that V_1 affected actual learning through stimulus-response asynchronism rather than through the strength of the eliciting stimulus. Hull was not clear on this point, but the final version of his theory does not seem to allow the deduction that amount learned increases with the intensity of the actual stimulus.

The only other theory of stimulus-intensity effects has been proposed by Perkins (1953) and Logan (1954). Working independently these contributors arrived at theories which are so similar that they will be treated as one for the present purpose. This theory is based on the assumption that, as a given response is being elicited by a given stimulus with reinforcement, so the stimulus gains excitatory properties which generalize to similar stimuli; conversely, the theory assumes that since the occurrence of the same response to other stimuli in the situation is not reinforced, then these other stimuli gain inhibitory properties which also generalize. The latter class of stimuli is said to be made up of "background cues" or intertrial stimuli, or to constitute the "contextual environment." This state of affairs, the result of differential conditioning, is represented schematically in Figure 1. If a tone is used as a conditioned stimulus (CS) in a classical conditioning situation, for example, then it is assumed to acquire the same degree of gross excitatory potential in separate groups trained with a weak tone (W in Case A on the left) and a strong tone (S in Case B on the right), and this generalizes along the stimulus-intensity dimension as shown. When a response is given to the intertrial stimulus of "silence" (0, the ambient noise level) during this training period, it is not reinforced and the intertrial stimulus acquires inhibitory potential which also generalizes. The further assumption that inhibition is to be subtracted from gross excitation to

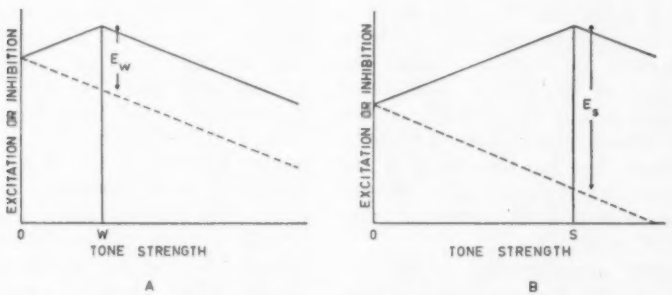

FIG. 1. Schematic diagram of the hypothetical gradients of excitation (continuous line) and inhibition (broken line) after reinforcement of responses to the onset of a weak tone (W) as CS (Case A) and a strong tone (S) as CS (Case B) with nonreinforcement of responses to the intertrial stimulus of "silence" (O). (Net excitation is greater for the strong CS—E_s—than for the weak CS—E_W.)

give net excitation allows the Perkins-Logan theory to predict greater net excitation for a strong tone (E_S) than for a weak tone (E_W), with the consequent expectation of a stronger response in the former case.

The Perkins-Logan theory assumes that amount learned is independent of the intensity of the eliciting stimulus. If it were to be shown that learning is superior with a stronger CS then the case for the deduction made above would be strengthened, gross excitation being greater with the strong than with the weak tone. A later deduction taking drive level into account, however, would be altered in this case. The present application of the theory also assumes that the intertrial or "background" stimulus acquires just enough inhibitory potential to prevent a response occurring between reinforced presentations of the stimulus; this should be the case, because as enough inhibition to block the response is generated so the response will no longer occur unreinforced and no more inhibition can accrue to the stimulus. It will also be noted that the generalization gradients have been depicted as linear, following Perkins (1953). This has been done in the present instance because it is felt that the exact shape is yet to be firmly established. Perhaps the weight of evidence favors the existence of negatively accelerated gradients (e.g., Raben, 1949) and one of the following deductions is made on this basis, but the remainder of the argument is not significantly affected by the assumption of a variety of other types of gradient.

Further consideration of the account of the Perkins-Logan theory given above, read in conjunction with Figure 1, may lead to the inference that stimulus-intensity effects are to be understood as the result of discrimination learning. In terms of the example depicted in Figure 1, a strong stimulus produces a stronger response not as a result of any inherent dynamism but rather because its presence is more readily discriminated from its absence. This parallel may be drawn in certain instances and will be considered wherever possible in the following discussion, but it is an inexact analogy in some respects and is therefore potentially misleading. Basically, the Perkins-Logan interpretation of the stimulus-intensity situation is put in terms of differential conditioning rather than traditional discrimination learning; i.e., the interest lies not in which of two or more responses will occur to a given stimulus, nor in which of two or more stimuli will elicit a given response, but in how strongly a given response will occur when elicited by a given stimulus. To put the matter another way, although the effect on a given response of variations in the strength of its eliciting stimulus depends upon generalized inhibition from the intertrial or background stimulus, these two stimuli are not necessarily presented simultaneously and no discrimination is therefore required. In the example cited above there is temporal or successive contrast between the positive CS and the negative intertrial stimuli. The case of a rat learning to run to a black card on a white background, with spatial or simultaneous contrast of "cue" and background stimuli, to which the Perkins-Logan theory has also been applied (e.g., Nygaard, 1958), resembles true discrimination learning more closely. It is still a form of spatial learning, however, for the cue and background stimuli hold constant positions in space and are not reversed in the one learning series for a given animal.

A final point of interest in the exposition of the differential conditioning theory is the probability that

it applies to variations in the *quality* as well as in the intensity of the stimulus. Predictable effects should ensue whenever the eliciting and background stimuli lie on some common qualitative dimension, with intensity held constant. There is no reason why experiments should not be conducted to test for stimulus-quality effects, but with the possibility considered, the present discussion will be limited to intensity effects.

STIMULUS INTENSITY AND LEARNING

The second of the two problems stated at the outset concerned the possibility that a response is better learned if the eliciting stimulus is strong during learning. A number of attempts have been made to investigate this possibility, but no satisfactory test has yet been devised, so that the problem remains unsolved and a challenge to future workers. For this reason the problem cannot be pursued at length now, although a brief account of its nature may clarify the discussion which follows.

The latest attack on the problem formed part of a study of eyelid conditioning conducted by Walker (1960), which may be used for illustrative purposes. Following the lead of Grant and Schneider (1948, 1949), Walker used a factorial design; she trained one group of subjects with a strong CS and another group with a weak CS, and then shifted half of each of these groups to weak and strong CSs, respectively, in extinction, leaving the other half of each group on the CS strength used in training. The nature of this experimental design is represented in Table 1, where the first letter in each cell stands for CS intensity during training, the second letter stands for CS intensity during extinction, and the measures are taken on the extinction series. Walker

TABLE 1

THE FACTORIAL DESIGN USED BY WALKER (1960)

Training CS intensity	Extinction CS intensity	
	Strong	Weak
Strong	SS	SW
Weak	WS	WW

sought to determine the effect of CS intensity on habit strength by comparing the performance in extinction of the two training groups, represented in the two rows of Table 1, but there proved to be no significant difference.

Unfortunately, this aspect of Walker's (1960) experiment is rendered invalid by the possible interfering effects of the shifts in CS intensity from training to extinction which occurred in half the subjects. The impact of these shifts may be treated in terms of both the Hullian and Perkins-Logan theories. Taking Hull's theory first, the shifts should bring changes both in generalized habit strength due to stimulus-intensity generalization and in stimulus-intensity dynamism (from V_1 to V_2). The former effect would always cause a loss of response strength whereas the latter factor could produce either an increase or a decrease therein. These two factors might obscure any effect of V_1 on ${}_sH_R$ and until some determination is made of their differential weights in contributing to ${}_sE_R$ no allowance can be made for them. It is thus impossible to infer from Walker's data that amount learned is independent of CS intensity. Walker did not make this inference; she suggested instead that extinction data may not provide a proper basis for testing the hypothesis but failed to give any reason for the suggestion.

Further reference to Figure 1 will

show how the Perkins-Logan theory also leads one to believe that shifts in CS intensity after training might obscure any effect on amount learned. On the basis of this theory the expected performance of Walker's (1960) four groups in extinction (Table 1) may be set down thus: WW—a continuous decline due to inhibition accruing to the CS as well as to the intertrial stimuli; WS—as for WW; SS—a continuous decline from a higher level of performance than that initially present in WW and WS; SW —a sharp decrease with the shift, from the same higher level, followed by a slower decline. If these differential effects are to be expected with a null hypothesis that CS intensity does not affect learning then it is clear that no alternative hypothesis may be subjected to a proper test. Recalling that Walker obtained no significant effects in extinction, with unconditioned stimulus (UCS) intensity held constant, it might seem that the predictions derived from the Perkins-Logan theory were not verified. It is doubtful, however, that Walker's analysis of the data did justice to the predictions, no account being taken, for example, of trends over successive trials in extinction. On the other hand it is conceivable that the discrepancy is due to the occurrence of superior learning with a stronger CS.

To return briefly to the analogy with discrimination learning, the shifts in CS intensity from training to test which are used in this type of study are reminiscent of the "transposition phenomenon." The analogy is still inexact, for the original form of transposition required the preservation of a relationship between two stimuli when the absolute intensity of both was changed; argument then took place as to whether the relationship was vital to the learner (gestalt theory) or to the theorist (S-R theory). In the factorial studies of stimulus-intensity effects, the absolute value of only the CS or cue stimulus is changed and that of the intertrial or background stimuli remains constant. Such experiments could be conducted, of course, with shifts in the intensity of both sources of stimulation, so as to approach true transposition more closely.

This short account of the possible role of stimulus intensity in learning, with Walker's (1960) study as an example of relevant experimental evidence, should suffice to lend weight to the claim that no satisfactory test of this aspect of the stimulus-intensity phenomenon has yet been devised. It is therefore fortunate that a clear distinction may be made between effects on learning and on performance. The remainder of this paper is accordingly devoted exclusively to effects on performance; it is hoped that the preceding examination of effects on learning will allow the reader to bear this important distinction in mind.

STIMULUS-OFFSET EFFECTS ON PERFORMANCE

In considering the implications of his theory, Logan (1954) suggested that the offset of a stimulus should be as effective as its onset in eliciting a response and that the strength of the response should vary with the intensity of the offset stimulus. It is difficult to see how a strict interpretation of Hull's stimulus-dynamism principle could predict such an outcome. Hansche and Grant (1960) have shown that offset and onset are equally effective as a CS in eyelid conditioning, and Schwartz (1958) found avoidance response strength in rats to vary directly with degree of change in CS intensity, regardless of

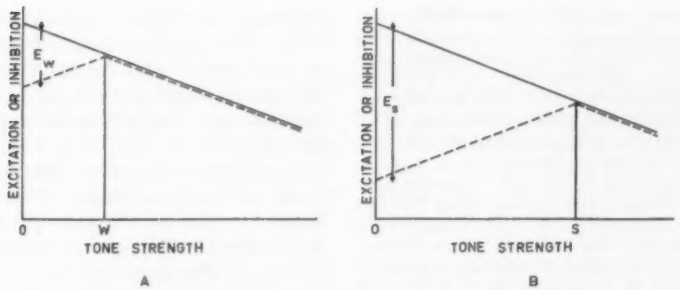

FIG. 2. Schematic diagram of the hypothetical gradients of excitation (continuous line) and inhibition (broken line) after reinforcement of responses to the offset of a tone (onset of "silence") as a CS with nonreinforcement of responses to a weak tone (Case A) or a strong tone (Case B) as the intertrial stimulus. (Net excitation is greater with the offset of the strong tone—E_s—than with the offset of the weak tone—E_W.)

the direction of the change. The detailed application of the Perkins-Logan theory to the case of stimulus offset as a CS is represented schematically in Figure 2, in which the symbols are the same as for Figure 1. In this situation it is the state of "silence" which should acquire excitatory properties, whereas the weak or strong intertrial stimuli come to inhibit the response. Even so, the net excitation with a strong intertrial stimulus (E_S) is still greater than that with a weak intertrial stimulus (E_W) after the excitation and inhibition have been built up in the course of training. Before any differential conditioning, on the other hand, there should be no difference in the strength of the responses elicited by the two degrees of offset. An experiment was conducted to test these deductions; it allowed the findings of Hansche and Grant, and of Schwartz, to be confirmed, but made wider generality of application possible by the use of the long-latency GSR.

EXPERIMENT I

The subjects were 24 male and 24 female volunteers from an introductory course in psychology at the University of Sydney. The intertrial stimulus was a 2000-cps sine-wave tone delivered to the subject through headphones at sound pressure levels of 80 db (strong) or 60 db (weak) and the CS was the offset of the tone for 2000 milliseconds. An attempt was made to choose sound levels such that the loud tone would not be so intense as to disturb the subject when presented almost continuously for 20 minutes and that the soft tone would not be so weak as to be below threshold for any subject. The UCS was a 90-cps spike-wave shock of 500-millisecond duration which was presented 500 milliseconds after the offset of the intertrial stimulus as CS. It was set at 2.27 milliamperes as measured with a 15,000-ohm resistor in place of the subject and was delivered through thimble-type electrodes of German silver .5 inch in length, attached to the third and fifth fingers of the right hand. The GSR was picked up with electrodes of a similar type 1 inch in length, attached to the second and fourth fingers of the right hand, through which was passed a constant direct current of 100 microamperes; the response was measured with a bridge-type circuit linked to an amplifier and ink-writing recorder. Conductive paste was used in conjunction with the shock and GSR electrodes. The duration of the CS, the UCS, and the CS-UCS interval was controlled by electronic timers. The subject was placed in a sound-reduced room completely separate from that containing the experimenter and the main apparatus.

The measure of performance used was the amplitude of the GSR on test trials (tone offset alone) which were interspersed with the training trials (tone offset followed by shock). Each series began with a test trial followed by one presentation of the UCS alone and a

second test trial. Then three further test trials were given after 1, 2, or 3 training trials (third test trial), after 4, 5, or 6 training trials (fourth test trial), and after 7, 8, or 9 training trials (fifth test trial). Within this arrangement six orders of trials were used, in which test trials were presented after the following numbers of training trials: 1, 5, 9; 1, 6, 8; 2, 4, 9; 2, 6, 7; 3, 4, 8; 3, 5, 7. Thus the mean number of preceding training trials for Test Trials 1 through 5 were 0, 0, 2, 5, and 8 trials successively. The subjects were trained either with the offset of the strong tone as the CS (Group S) or with the offset of the weak tone as the CS (Group W). They were allocated to the two groups of the experiment and to the six orders of trials in a counterbalanced sequence imposed upon the order in which they appeared at the laboratory with the restriction that there be 12 men and 12 women in each of the two groups. The intertrial interval was at least 60 seconds and depended in part on the time taken to achieve a steady level of skin resistance; it rarely exceeded 75 seconds and averaged about 65 seconds.

The performance curves derived from the data are shown in Figure 3, in which the first point on the two curves is based on each subject's mean response over the first and second test trials. The performance of Group S was compared with that of Group W before any training trials and a *U* test showed that the difference was not significant ($z = 1.51$, $p = .13$). The nonparametric test was preferred because of some evidence of nonnormality in the distribution of scores. To ensure that the groups were comparable at this stage a *U* test was also applied to the responses on the second trial, consisting of the UCS (shock) alone. The mean values were 1.57 and 1.74 micromhos for Groups S and W, respectively, and the difference was not significant ($z = 0.22$, $p = .41$). Inspection of the results showed that Group S improved more than Group W in conditioning and the significance of this finding was tested by taking the difference between the amplitude of the response on the second and third test trials for each subject. A *U* test showed that the gain was significantly greater for Group S ($z = 2.12$, $p = .03$).

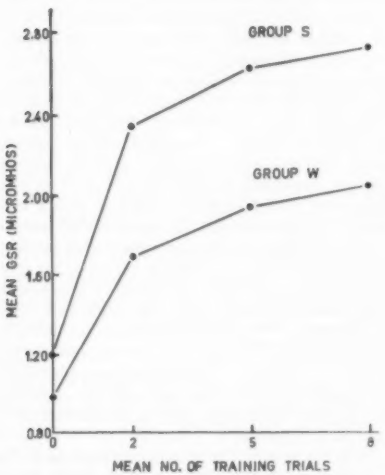

FIG. 3. Response strength on test trials in Experiment I as a function of the mean number of preceding training trials for Groups S (strong tone off) and W (weak tone off).

The results of this experiment confirmed that the offset of a stimulus may be used as a CS and that, in the course of acquisition, the offset of a strong CS comes to elicit a stronger response than the offset of a weak CS, as the Perkins-Logan theory predicts. It is emphasized that this effect is being treated as a matter of *performance* only (Figure 2), it is not being inferred that true *learning* is greater with the offset of the stronger CS. Hull's principle of stimulus-intensity dynamism was couched purely in terms of the absolute strength of the eliciting stimulus, acting through its onset. In formulating the principle, he cited evidence that rats jump faster to white than to black cards, but a law of absolute stimulus strengths is manifestly inapplicable to the case where animals respond more strongly to a

black (low intensity) card on a white background than to a white (high intensity) card on a white background (Nygaard, 1958). It is exactly an effect such as this, on the contrary, which is predicted by the Perkins-Logan theory. A further difficulty in the application of Hull's theory to the present data is its limitation to short-latency responses and its failure therefore to explain such effects as have been obtained with the GSR. The two main results of Experiment I were the demonstration that "silence" is comparable in effectiveness with an audible tone, as a CS, and that the response elicited by a brief period of "silence" is stronger with a loud tone as the intertrial stimulus than with a weak tone between trials. The former result is comparable with the common finding in discrimination learning studies that, with a white card and a black card presented simultaneously, for example, either may be made the "positive" cue in the sense that approach to it is followed by reinforcement, provided allowance is made for initial position preferences. The second result of Experiment I is presumably analogous with the equally common finding that rate of discrimination learning is directly proportional to the degree of difference between the positive and negative cues, for example, that learning proceeds more quickly with white and black cards as the cues than with white and gray or with black and gray.

The results of Experiment I suggest a further prediction about generalization effects which follows the application of the Perkins-Logan theory to the stimulus-offset situation. It is first necessary, however, to deal with these effects as they occur in the more usual case of stimulus onset (Figure 1). The clearest evidence on stimulus-intensity generalization has probably been provided by Grice and Saltz (1950), who found asymmetrical gradients falling rapidly with test stimuli weaker than the training stimulus and more slowly for stronger test stimuli. Examination of Figure 1 shows that, according to the present formulation of the Perkins-Logan theory, tests with stronger stimuli should not yield a weaker response, as net excitatory potential stays constant. The reason why this deduction does not correspond with the facts may well lie in the assumption of linear gradients of generalized excitation and inhibition. If negatively accelerated gradients are substituted for the linear gradients of Figure 1 then tests with stronger stimuli should also produce a weaker response. The gradients of *net* excitatory potential should still be asymmetrical; Hull (1949) predicted this asymmetry by allowing for stimulus dynamism, but the substitution of negatively accelerated for linear gradients allows the Perkins-Logan theory to predict the same outcome. According to the latter theory, however, when the eliciting and background stimuli do not lie on any common dimension, then the gradients of *net* excitatory potential should be symmetrical about the value of the eliciting stimulus. This lack of communality and the consequent symmetry are more likely with qualitative differences than with changes in stimulus intensity.

The application of these considerations to the case of stimulus offset gives rise to the prediction which is still untested. When the ambient noise level ("silence") of, say, 50 db is used as the CS then tests of stimulus generalization may be made by briefly changing the intensity of the intertrial stimulus to some value other than 50 db. For example, the intertrial stimulus could be switched

from 80 db to 90 db or to 70 db for 2000 milliseconds (see Case B of Figure 2). *With negatively accelerated gradients assumed*, a momentary increase or decrease in the intensity of the background stimulus should produce a response which is always weaker than the original CR but relatively stronger with the decrease than with the increase. Momentary decreases in the intensity of the intertrial stimulus *below* the intensity of the CS (50 db), however, should produce another gradient of net excitation. Thus a plot of the curve relating net excitatory potential to stimulus intensity should show two points of inflection, one at the value of the CS and another at the value of the intertrial stimulus. This effect would be easier to detect if the CS as well as the intertrial stimulus was set at some intensity above the ambient noise level, at 60 db, for example. Inasmuch as the more usual stimulus-onset situation involves an intertrial stimulus of some intensity, however weak, then the same effect should be found there, but its presence would be more difficult to detect for technical reasons.

MODIFIED HULLIAN THEORY

The results of Experiment I showed that a strict, literal interpretation of Hull's principle of stimulus-intensity dynamism must be abandoned, but they also suggest a modification by which Hull's approach to the problem might be sustained. Along with all other S-R theorists Hull assumed that a stimulus acts through its occurrence or presence and this is the source of difficulty in dealing with the cessation or "absence" of a stimulus as an active factor. The difficulty could be overcome with the supposition that a stimulus acts through any *change* in the intensity of the physical energies impinging upon the organism's receptors and that the effectiveness of the stimulation varies with the degree of change, *regardless of its direction*. An assumption of this kind is supported by physiological evidence for the existence of nerve fibers which respond only to changes in stimulation (e.g., Granit, 1947). Thus it is possible to contrast a neo-Hullian or modified stimulus-intensity dynamism principle with the differential conditioning theory of Perkins and Logan, and a second experiment was designed with this aim in view. In its revised form the stimulus-intensity dynamism principle should ascribe equal effectiveness to the onset and the offset of a stimulus, provided that degree of change is held constant, and should not lead one to expect any effect with a shift from onset to offset (or vice versa) as the stimulus over a series of trials with the same subject. According to the differential conditioning theory, on the other hand, once a given stimulus has acquired inhibitory properties for a given subject then negative transfer should occur if the same stimulus is then used to elicit the response in the same subject, as happens with a shift from onset to offset or vice versa. In order that due allowance might be made for the limitation of the Hullian formulation to short-latency responses, this experiment was conducted with a key pressing movement instead of the GSR.

EXPERIMENT II

The subjects were 12 members of undergraduate courses at the University of Sydney. The form of physical energy manipulated as an eliciting stimulus was the same tone as in Experiment I, again delivered through headphones at 80 db or 60 db SPL. The subject was placed in the sound-reduced room and asked to respond by pressing a telegraph key fitted with a small platform upon which rested the fingers of the right hand. Reaction times were measured with an electronic chronoscope. Each subject was presented with the four combinations of stimulus condi-

tions in blocks of nine trials, viz., weak tone offset (A), strong tone offset (B), weak tone onset (C), and strong tone onset (D). With subjects numbered in order of testing, they underwent the four conditions in one session in the following orders: Subjects 1, 5, and 9—CDAB; Subjects 2, 6, and 10—DCBA; Subjects 3, 7, and 11—ABCD; Subjects 4, 8, and 12—BADC. Before each block of trials the subject was told what the nature of the signal was to be and was simply asked to press the key as quickly as possible upon hearing the signal. A verbal ready signal was given over an intercom approximately 2 seconds before each stimulus signal and the duration of the occurrence or absence of the tone was 2 seconds. The interval between trials was approximately 15 seconds.

The design of the experiment was such that the scores of the 12 subjects could be averaged for each trial irrespective of the conditions under which they were tested, so as to cancel out the main effects due to individual differences and to amount and direction of change in stimulus intensity as the signal. The adoption of this procedure left not only practice as a main effect but also *shifts* in the nature of the signal. Between Trials 9–10, 18–19, and 27–28 there were shifts in *degree* of change in stimulus intensity (small to large or large to small), but between Trials 18–19 there occurred a shift in *direction* of change in stimulus intensity (onset to offset or offset to onset) as well as in degree. Mean performance for the ordinal series of trials is depicted in Figure 4 and it is there clear that whereas the shifts in degree of change simply merged with the general practice effect, the shift in direction as well as degree caused considerable negative transfer. Wilcoxon's rank test was applied to a comparison of the scores on Trials 18 and 19 and this showed that the negative transfer was significant ($z = 3.43$, $p < .001$). These results are interpreted as clearly supporting the Perkins-Logan theory as against a modified stimulus-dynamism postulate.

The findings of the first experiment and other studies were also confirmed in the data. Working with the total reaction time over the last five trials in the block of nine for each subject, in order to reduce the practice effects, rank tests were used to assess the main effects of weak versus strong stimuli and of onset versus offset as a signal. The former effect was significant ($z = 3.19$, $p < .01$) but the

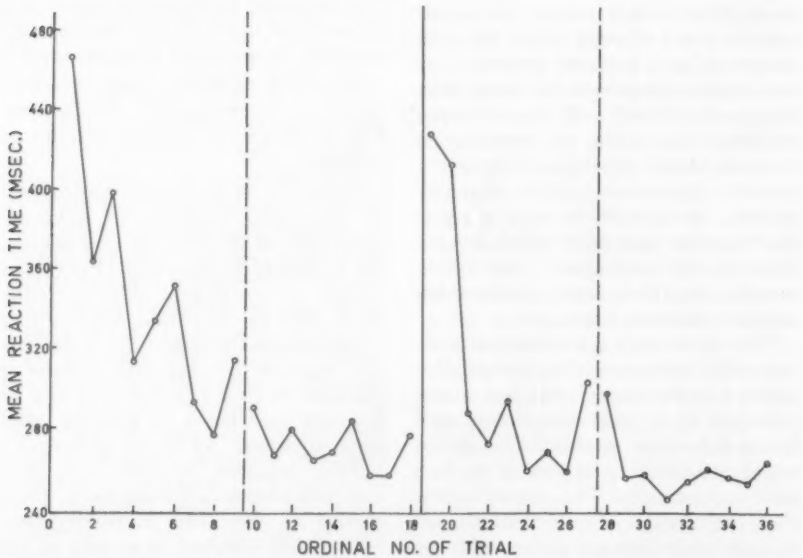

FIG. 4. Performance in Experiment II as a function of the ordinal number of the trial regardless of condition of testing. (Shifts in degree of change in stimulus intensity as the signal occurred between Trials 9–10, 18–19, and 27–28, but between Trials 18–19 there was also a shift in direction of change.)

latter effect was not significant ($z = 1.55$, $p = .12$). Thus, as Schwartz (1958) observed in the case of rats, strength of response varied with amount of change in stimulus intensity, regardless of the direction of the change.

The results of the second experiment nullified the attempt to modify Hull's treatment of the stimulus-intensity phenomenon and lent further weight to the Perkins-Logan theory. Apart from the direct disproof of the modified stimulus-dynamism principle, the differential conditioning theory should be preferred as the more parsimonious, for it avoids the assumption that a stimulus acts through change in intensity rather than its mere presence. This is not to deny rapid adaptation, so that a stimulus is most effective at the moment of onset, but it does imply that mere offset cannot be regarded as having stimulus properties. In terms of the Perkins-Logan theory, offset is effective because it inevitably means the simultaneous onset of some other physical energy as far as receptor processes are concerned. It should be noted that the terms "offset" and "onset" need not imply any direct intervention on the part of the experiment, as in the present experiments, but may be applied to changes in the energies affecting the receptors which are induced by the organism's own movements, as in instrumental conditioning or discrimination learning.

The use of such a precise and well-controlled response as key pressing by humans in this experiment has drawn attention to a point which may well demand further study if the differential conditioning theory is to find wide acceptance. The theory states that background or intertrial stimuli acquire inhibitory properties with respect to a given response when that response occurs in their presence without reinforcement. Although it was observed in Experiment I that GSRs commonly occurred between trials early in training, there were relatively few cases of key pressing to the intertrial stimulus in Experiment II. Some subjects who did respond before the signal of tone onset or offset apologized to the experimenter or acted as if they had behaved foolishly; the intertrial GSRs, on the other hand, would have passed undetected by the subject. Thus it may be necessary to assume that with such "voluntary" responses as key pressing it is enough for the response to occur implicitly without reinforcement for inhibition to accrue to the stimulus present at the time. This raises the further problem of the nature of the reinforcement for learning to give a "voluntary" response more and more quickly under instruction. No attempt will be made to enter into a discussion of such issues here, but it is clear from Figure 4 that learning did occur in a situation of this type and the immediate drop in response strength with a reversal of signaling and intertrial stimuli is in keeping with the Perkins-Logan theory as it is formulated at present.

Referring again to the comparison of the stimulus-intensity situation with that of discrimination learning, the results of Experiment II permit a parallel to be drawn. This is to the effect that in discrimination learning with white as the positive and black as the negative cue, negative transfer is obtained with a reversal to white as negative and black as positive after some training. In terms of Spence's (1937) analysis of the transposition phenomenon, furthermore, there should be some loss in performance with a shift from white to gray as the positive cue, with black negative throughout, or from black to gray as the positive cue with white constantly

negative, because the difference between the positive and negative cues is less after the shifts; conversely, there should be gains in performance with shifts of the positive cue in the reverse direction, i.e., from gray to white with black negative or from gray to black with white negative, there being a greater difference between the positive and negative cues after the shifts. Comparable with these findings or expectations are the facts that the reversal of signaling and intertrial stimuli in Experiment II led to a marked drop in performance, and that level of performance depended on the degree of difference in intensity between these two stimuli. The latter feature of Experiment II did not show up in Figure 4, of course, because the design counterbalanced large and small intensity differences at successive shifts, but it was clearly evident when degree of change was studied as a main effect, as in Experiment I also.

STIMULUS INTENSITY AND DRIVE STRENGTH

The third and fourth studies of this series were conducted with a common aim, the determination of the general nature of the joint effect upon response strength of variations in stimulus intensity and drive strength. As noted at the outset, Hull *assumed* a multiplicative relationship between stimulus-intensity dynamism (V) and drive strength (D), but the Perkins-Logan theory should allow a *deduction* to be made in this respect. If the excitation accruing to the eliciting stimulus and the negative effects of generalized inhibition are set down in a conventional way, then the determination of the strength of the response in a variety of situations may be represented as

$$R = f(E) = f(H \times D - I)$$

where E is net excitatory potential, H is habit strength, D is drive strength, and I is inhibition (Spence, 1960). The application of this theorizing to the case of a response which is elicited by both weak and strong values of the stimulus, with H held constant, leads to the deduction that variations in the strength of the eliciting stimulus (i.e., in degree of inhibition) will act *additively* with variations in D. It is this deduction from the Perkins-Logan theory, opposed as it is to Hull's assumption, which was tested in the following two experiments. Because of its greater sensitivity to motivational effects and its "involuntary" nature, use was again made of the GSR. A factorial design was used to test for the additive or multiplicative relationship, it being assumed that the latter would be reflected in significant interaction between the variables.

Instead of using stimulus offset as the CS in a situation requiring new learning (as in Experiment I), capital was made of the fact that, due either to transfer from past experience or to the nature of the innate reflex, most subjects give a GSR to a tone without formal training, the response to a strong tone being greater than that to a weak tone. While many subjects also give GSRs between trials at first, the intertrial responses soon fall to a minimum frequency. The experiment therefore proceeded straight to an examination of the relation between this existing stimulus-intensity effect and drive strength. Variation in drive strength was achieved through the strength of shocks administered at random among the test presentations of the weak and strong tones, a technique used in a number of recent studies for this purpose (e.g., Spence, Haggard, & Ross, 1958). The repeated presentation of a tone as an eliciting stimulus for the GSR without

reinforcement, however, produces the decrement in performance usually ascribed to either or both of the processes of adaptation and experimental extinction. In order to carry out a sufficient number of test trials (tone alone) it was necessary in one of the two experiments to interpolate training sessions among test sessions. The training sessions consisted of regular conditioning trials, with tone as CS and shock as UCS. The separate test series in this case each required an appropriate combination of tone strength and shock strength, but care was taken in the training series to ensure that each "boosting" session included all combinations of CS and UCS intensity in order that no artifacts would be introduced by the conditioning of the GSR to strength of tone as such or of the size of the response as such. By this means the assumption of constant habit strength for the GSR to the tone was justified.

The detail of the deduction to be tested under these conditions is as follows. Let E_{WW}, E_{WS}, E_{SW}, and E_{SS} represent the degree of net excitatory potential generated with the combinations of weak tone and weak shock, weak tone and strong shock, strong tone and weak shock, and strong tone and strong shock, respectively. Then

$$E_{WW} = H \times D- - \bar{I}+$$
$$E_{WS} = H \times D+ - \bar{I}+$$
$$E_{SW} = H \times D- - \bar{I}-$$

and

$$E_{SS} = H \times D+ - \bar{I}-$$

where H is constant, $D+$ and $D-$ are due to strong and weak shock, respectively, and $\bar{I}+$ and $\bar{I}-$ are the values of the generalized inhibition present with weak and strong tones, respectively, as eliciting stimuli. This state of affairs is represented in Figure 5.

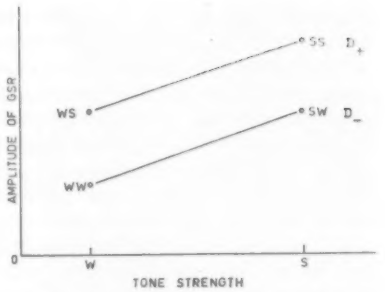

FIG. 5. The application of the Perkins-Logan theory to the covariation of stimulus strength (tone) and drive strength (shock). (The broken line indicates the gradient of inhibition, originating at the intertrial stimulus—O—and generalizing to the weak tone—$\bar{I}+$—and strong tone—$\bar{I}-$. The E values are for the four combinations of weak or strong tone—first subscript—with weak or strong shock—second subscript.)

The use of a factorial design allowed a test of an additive relationship in the absence of interaction between the variables. As set out in Figure 6, this test may be made by comparing the difference between the mean GSRs for strong and weak shock with a weak tone (WS-WW in Figure 6) with the difference between means for strong and weak shocks with a strong tone (SS-SW). The comparison may

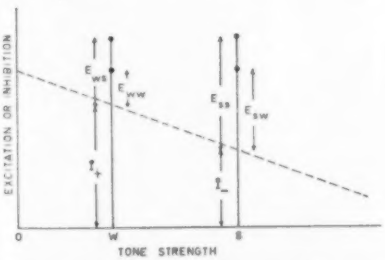

FIG. 6. Schematic representation of an additive relationship between stimulus strength (tone) and drive strength (D) generated through shock. (The letters indicate the four combinations of weak or strong tone—first letter—with weak or strong shock—second letter.)

also be made between the differences for strong and weak tone with a weak shock (SW-WW) and for strong and weak tone with a strong shock (SS-WS). The specific predictions from the Perkins-Logan theory are obtained thus:

$$E_{WS} - E_{WW} = HD+ - \bar{I}+ - HD- + \bar{I}+ = H(D+ - D-)$$

and

$$E_{SS} - E_{SW} = HD+ - \bar{I}- - HD- + \bar{I}- = H(D+ - D-)$$

when it is to be expected that the differences will be equal and that there will be an additive relationship such as is depicted in Figure 6. It also follows that both $E_{SW} - E_{WW}$ and $E_{SS} - E_{WS}$ are equal to $\bar{I}+ - \bar{I}-$.

EXPERIMENT III

The subjects were one man and one woman. The apparatus differed from that used in Experiment I in the following respects. The eliciting stimulus was a 2000-millisecond, 2000-cps sine-wave tone delivered to the subject through headphones at 60 db (weak tone), 75 db (medium tone), or 90 db (strong tone) SPL. The noxious stimulus used to induce drive was a 500-millisecond, 90-cps, spike-wave shock delivered through copper electrodes in the form of spring clamps $\frac{3}{4}$ inch wide which almost encircled the third and fifth fingers of the right hand near the fingertips, and it was set at 2.27 (weak shock), 3.20 (medium shock), or 4.80 (strong shock) milliamperes. After being placed in the apparatus the subject was merely told to keep still and warned of the first shock and the first tone. Both subjects underwent a number of test sessions at the rate of one each day at the same hour (11:00 A.M.–12:30 P.M.) and each session consisted of two shock trials, six tone and six shock trials in random order, and one tone trial. Shock strength and tone strength were held constant throughout any one test session but were varied from session to session in accordance with the basic aim of the experiment. The test sessions for the female subject involved the six combinations of weak (W), medium (M), and strong (S) tone with weak (W) and medium (M) shock, and the order of combinations over sessions was WW (weak tone and weak shock), SM, WM, SW, MW, and MM. This order did not seem to allow any biasing effect due to adaptation and if anything appeared to favor the occurrence of a direct multiplicative relationship. The test sessions for the male subject involved the four combinations of weak and strong tone with weak and strong shock. In this case it was possible to use the more desirable ABBA sequence and the combinations were used in the order SS, SW, WS, WW, WW, WS, SW, and SS. Twenty-four hours before each test session (following the previous test session) the subject was given a training series of nine conditioning trials (tone-shock with a 500-millisecond interstimulus interval) to combat adaptation and extinction effects, and each training series included all the combinations of tone strength and shock strength.

The data for each subject were treated separately and they are presented in Table 2. The results of an analysis of variance are set out in Table 3. In neither case did the tone-shock interaction, to be expected with a multiplicative relationship, prove statistically significant.

EXPERIMENT IV

The subjects were 20 volunteers (9 men and 11 women) from an introductory psychology course at the University of Sydney. With two exceptions the apparatus and stimuli were identical with those used in Experiment III. The copper shock electrodes were replaced with a type similar to those used to pick up the GSR, and only the 1.60-milliampere and 3.20-milliampere shocks were used, being des-

TABLE 2

MEAN GSR TO TONE (micromhos) IN EXPERIMENT III

Subject	Tone strength	Shock strength		
		Strong	Medium	Weak
Female	Strong	—	4.79	2.60
	Medium	—	2.74	1.26
	Weak	—	2.91	0.16
Male	Strong	1.30		0.27
	Weak	0.52	—	0.07

TABLE 3

SUMMARY OF ANALYSIS OF VARIANCE OF GSRs IN EXPERIMENT III

Subject	Source	*df*	*MS*	*F*
Female	Tone (A)	2	1797	3.24
	Shock (B)	1	4864	8.78**
	A × B	2	139	0.25
	Residual	36	554	
Male	Tone (A)	1	340	4.05*
	Shock (B)	1	757	9.01**
	A × B	1	109	1.30
	Residual	52	84	

* $p < .05$.
** $p < .01$.

ignated weak and strong, respectively, henceforth. The measures for each subject were taken in a series consisting of one medium tone trial, two shock trials, and, in a different random order for each subject, five weak tone, five strong tone, and eight shock trials. The subjects were matched in pairs on the basis of the amplitude of the GSR on the first tone trial (medium), one member of each pair then being tested with strong neighboring shock and the other with weak shock. Thus whereas shock strength was constant during the session for any one subject, both strong tone and weak tone trials were given to all subjects. Because each subject was tested in one session, with only 11 tone alone trials, adaptation and extinction effects were relatively slight and no training trials were used. Once again use was made of the facts that subjects enter this situation with a tendency to respond to the tone and that intertrial responses are soon inhibited. In view of the matching of subjects across shock strengths the data were treated as if the same subjects had undergone each combination of stimulus strengths and the results of an analysis of variance are shown in Table 4. Again the tone-shock interaction was not significant, as might be expected from an inspection of the data in Table 5.

The data of Experiments III and IV confirmed the deduction from the Perkins-Logan theory and were incompatible with the Hullian assumption of a multiplicative relationship between stimulus strength and drive strength. Another aspect of Walker's (1960) study, cited earlier, produced

TABLE 4

SUMMARY OF ANALYSIS OF VARIANCE OF GSRs IN EXPERIMENT IV

Source	*df*	*MS*	*F*
Tone (A)	1	2220	42.69**
Shock (B)	1	152	2.92
Subjects (C)	9	243	4.67*
A × B	1	2	0.04
A × C	9	276	5.31*
B × C	9	233	4.52*
A × B × C	9	52	

* $p < .05$.
** $p < .01$.

a similar finding. Half the subjects were trained with a weak UCS (air puff) and half with a strong UCS, and if this noxious UCS is thought of as a source of drive then her experiment also allowed a test of the relationship between stimulus intensity and drive strength. Walker found significant differences in acquisition due to both CS intensity and UCS intensity, but there was no interaction between these variables. This result cannot be taken as clear-cut, for the superior reinforcement of a strong UCS may have produced superior learning (Spence et al., 1958), as the strong CS may also have done. If either of these latter factors had produced differential habit strengths then a multiplicative relationship should have been obtained; the reason for this expectation emerges from a comparison with discrimination learning which follows shortly.

TABLE 5

MEAN GSR TO TONE (micromhos) IN EXPERIMENT IV

Tone strength	Shock strength	
	Strong	Weak
Strong	2.51	2.08
Weak	0.98	0.63

At least two other studies have produced evidence for Hull's formulation and some reconciliation must be made of these findings with the present data. First, with reference to an experiment on eyelid conditioning, Spence (1956) has reported a significant interaction between CS strength and anxiety level at asymptotic values of response strength. Second, Castaneda (1956) found the same type of relationship to hold for the latency and amplitude of a manual response, again with manifest anxiety as the drive variable. A common property of these two studies was the use of anxiety as a means of manipulating drive strength and the particular form of the data might be explained in terms of the "reactive" hypothesis of anxiety, i.e., with the idea that the difference between high and low anxious subjects from the motivational point of view is only apparent when they are presented with a "strong" eliciting stimulus or, in terms of the Perkins-Logan theory, with a stimulus which evokes little generalized inhibition. According to this view the multiplicative effect is obtained with anxiety as a drive variable because the effective D for the high anxious subjects is greater with a strong than with a weak eliciting stimulus. In the present experiments, on the other hand, the after effects of shock (D) are the same for strong and weak eliciting stimuli (tones), the reactive hypothesis does not apply, and the relationship under study takes the simpler additive form. It should also be noted that the experiment quoted by Spence was also complicated by the possibility that true learning was superior with the stronger CS.

It may be suggested that the rationale of Experiments III and IV was invalidated by the use of the GSR. In Experiment I it was found that effects predicted by the differential conditioning theory could be obtained with the GSR, so there can be no objection to its long latency as such, but a further difficulty may be seen in the fact that the GSR has often been used as a measure of emotion. Spence (1956) has assumed that an emotional response of an implicit nature (r_e) underlies the effects of aversive motivation (D) as generated by shocks, and the GSR might be taken as reflecting the strength of this emotional activity. If this argument be extended to allow that any stimulus, including a tone, arouses some emotional activity, then the use of the GSR in these experiments on the combined action of the tone and the aftereffects of shock might simply reflect the summation of two forms of emotional activity. There are at least two answers to this objection. In the first place, no theorist has supposed that a weak tone is motivating and it should follow that the granting of motivational properties to a strong tone only leads to the expectation of a multiplicative relationship between tone strength and shock strength. Second, it is likely that the latency of the implicit emotional or fear response is too long to affect even the relatively long-latency GSR. For the present purposes, therefore, it seems appropriate to treat the GSR as any other response within the framework of the differential conditioning theory, but to capitalize on its long latency in the manner indicated earlier.

Another point which must be considered is the methodological criticism recently levelled by Miller (1959) at the present type of study, which seeks the form of a relationship by simply taking representative values instead of covering the entire range of both variables. Miller's point is that the form of the relationship may vary with the location of the sampled

values within the entire range, and this is true if one of the variables is the number of reinforced trials (N), which causes response strength to increase in the form of a negative growth function or some kind of nonlinear progression. In the present instance, however, both D and $\bar{I}$ are assumed to bear a linear relationship to shock strength and tone strength, respectively, and it should not matter what values are chosen for experimentation within limits set by the introduction of physical damage factors with extremely strong shocks and loud tones. Even if a negatively accelerated relationship were assumed to hold between $\bar{I}$ and tone strength, as is more likely, then the deduction of an additive effect made above should still be valid, for the differences in E values for weak and strong shock at various tone strengths are not affected by the nature of this relationship and should stay constant (Figure 5). Quite apart from this argument it is felt that the values used in the present experiments covered the range to a reasonable degree; as Table 2 shows, the weak tone-weak shock combination was so weak as to produce a barely measurable response in the male subject on most trials, the mean value being .07 micromhos, whereas the strong tone-medium shock combination produced large responses with a mean of 4.79 micromhos in the female subject.

A second methodological factor which might affect the validity of any attempt to determine the nature of the joint effect of two variables in a theoretical system is the unit of measurement to be used for the dependent variable. The deduction from the Perkins-Logan theory of an additive relationship between stimulus strength and drive strength (independent variables) as joint determiners of net excitatory potential (intervening variable) was tested through the amplitude of the GSR (dependent variable). The link between net excitatory potential (E) and amplitude of GSR thus becomes crucial if the additive effect deduced to apply to the former variable is to be reflected in the latter variable. It is fortunate, therefore, that evidence of a simple linear relationship between these two variables is available (Spence, 1956, pp. 113–116). The nature of the scales for quantifying H, D, and I need not enter into the question, for the expression in these terms distinguishing response strength at the two drive levels is identical for weak and strong eliciting stimuli. While amplitude of response may therefore be used with confidence as a measure in the test of the deduction from the differential conditioning theory, other measures of response strength may not be equally appropriate. The present state of behavior theory leads to the deduction that frequency of response is a normal ogival function of superthreshold excitatory potential, in which case the evidence of Spence (1956) and Walker (1960) with frequency of conditioned response (CR) as a measure of E may be suspect. Latency of response should theoretically bear the same relationship to E as amplitude as a measure, so that the data of Castaneda (1956) which were also cited above are less open to this criticism.

Further comparisons may be made with discrimination learning. In deriving the prediction of an additive relationship from the Perkins-Logan theory, use was made of the formulation

$$R = f(E) = f(H \times D - I).$$

It will be noted that D is assumed only to multiply H and not I. This is a long-standing assumption, used by both Hull and Spence, and it is

supported by a wide variety of data. In the usual discrimination learning situation the learner is presented with two cue stimuli on each trial, with the response to one cue reinforced and to the other not reinforced. In these circumstances the positive cue should acquire habit strength (H) which generalizes to the negative cue ($\bar{H}$), while the negative cue acquires inhibitory potential (I) which generalizes to the positive cue ($\bar{I}$). If some uniform background is present throughout the experiment, and this background stimulus lies on the same dimension as the positive and negative stimuli, then it too will acquire inhibitory properties (I_B) which generalize to the positive stimulus ($\bar{I}_{B+}$) and to the negative stimulus ($\bar{I}_{B-}$). The application of the formula to give the net excitatory potential of the positive stimulus ($E+$) and the negative stimulus ($E-$) then gives the following expressions:

$$E+ = H \times D - (\bar{I} + \bar{I}_{B+})$$

and

$$E- = \bar{H} \times D - (I + \bar{I}_{B-})$$

when it follows that, if the two E values under high drive ($D+$) be represented as $E++$ and $E-+$, and under low drive ($D-$) as $E+-$ and $E--$,

$$E++ - E-+ = D+(H - \bar{H}) - k$$

and

$$E+- - E-- = D-(H - \bar{H}) - k$$

where k is a constant representing the total of the four sources of inhibition. Now $D+$ is greater than $D-$, so it has been deduced that discrimination learning under the stipulated conditions should be superior with high drive. The other derivation analogous with that made for the stimulus intensity situation requires a comparison of the E values for the positive stimulus with high and low drive ($E++ - E+-$) and the negative stimulus with high and low drive ($E-+ - E--$), as follows:

$$E++ - E+- = H(D+ - D-)$$

and

$$E-+ - E-- = \bar{H}(D+ - D-)$$

and since H is greater than $\bar{H}$, because the latter has generalized from the former, it is to be expected that there will be interaction between the degree of separation of the stimulus cues on the one hand and drive level during training, on the other hand. Reference back to the deductions about the joint effects of stimulus intensity and drive strength which were tested in Experiments III and IV will show that the assumption of constant habit strength is there crucial, for any difference in this respect between strong and weak eliciting stimuli will lead to a multiplicative rather than an additive relationship with drive strength. Of equal importance is the limitation of drive effects to habit strength and the supposition that drive does not multiply inhibition; modification of the basic formula to read

$$R = f(E) = f(D \times H - D \times I)$$

would also lead to the prediction of a multiplicative effect.

In a recent review of drive effects in spatial discrimination learning, Brown (1961) has reached a conclusion which tallies exactly with the deductions made above, i.e., that when the habit strengths of the response tendencies for the positive and negative cues are held equal, drive differences do not affect percentage choice of the correct response; if a correction method of learning is used, or if the two responses are reinforced with differential fre-

quencies, then drive effects become apparent. Brown notes, however, that there are too few relevant data to allow any such firm conclusion in the case of nonspatial discrimination learning. If anything in the latter case, performance improves with increased motivation, and this may be due to the greater difficulty in keeping habit strengths equal in the nonspatial situation with forced trials. It should also be noted that many of the studies cited by Brown involve instrumental approach learning, and that under these conditions the mere occurrence of the response, with or without reward, is assumed to produce an increment of habit strength; for this reason, forced runs to the nonrewarded side of a T maze, for example, are thought to produce as much learning as a free run to the rewarded side, and the increasing preference for the rewarded side is explained in terms of incentive motivation (Spence, 1956). It is therefore as well to emphasize the earlier caution that the analogy between the mechanism of the Perkins-Logan theory and discrimination learning is inexact, and that stimulus-intensity effects are more safely treated as due to differential classical conditioning.

Before concluding this discussion of the interrelated effects of stimulus intensity and drive strength, attention must be drawn to some recent comments by the author on drive effects in conflict (Champion, 1961). In order to attempt a resolution of an apparent inconsistency in the S-R treatment of conflict, it was assumed that changes in the strength of separate drives, such as hunger and fear, not only produce a common result by contributing to overall drive level, according to the generalized drive hypothesis, but also lead to distinguishable effects of a selective nature through the dynamism of the drive stimuli. Thus it was supposed that an increase in the hunger of a rat in approach-avoidance conflict affects the generalized habit strength responsible for the gradients of both approach and avoidance (generalized drive) and also acts selectively on the approach gradient through the strength of the hunger-drive stimulus. Although the evidence presented in the present account of stimulus-intensity effects would seem to force one to discard the *assumption* of a stimulus-dynamism principle, the *facts* of stimulus-intensity changes remain unchallenged and it is still possible to hypothesize that a stronger drive stimulus elicits a stronger response, other things being equal. As for the *explanation* of such a process, the substitution of the dimension of drive strength for that of the intensity of the eliciting external stimulus may pose further problems. It would have to be assumed that the response under study is not reinforced when the organism is satiated, so that satiation stimuli take on inhibitory properties. This latter inference is similar to some recent proposals by Estes (1958). More generally, however, the present line of argument preserves the purely motivational or energizing property of drives, while allowing them stimulus properties as well, whereas Estes appears to be moving towards a purely associative explanation of drive effects.

An experimental test of these alternatives might be conducted along the following lines. Suppose that a group of rats was trained to perform some act such as lever pressing at 4 hours food deprivation; the use of some reinforcement other than food would make this learning technically easier to obtain. Suppose also that the same animals were regularly placed in the apparatus at 12 hours deprivation but that the response of

lever pressing was never reinforced at this level. Considering only the role of the drive stimuli and assuming that the effect of drive is purely associative, the application of the Perkins-Logan theory to this form of differential training is set out in Figure 7. This figure differs from preceding representations in that habit strength rather than gross excitatory potential is plotted on the ordinate, along with inhibition. The lever pressing response is reinforced at 4 hours deprivation, and 6 units of habit strength are arbitrarily assigned to the drive stimuli at that level ($H = 6$); because of drive stimulus generalization, 5 and 4 units of generalized habit strength ($\bar{H}$) are assigned to the drive stimuli at 8 and 12 hours deprivation, respectively. The response is also allowed to occur at 12 hours deprivation, but it is never reinforced at that level, so that enough inhibition (I) accumulates to cancel the generalized habit strength, viz., 4 units; through drive stimulus generalization again, 3 and 2 units of generalized inhibition ($\bar{I}$) are assigned to the drive stimuli at 8 and 4 hours deprivation, respectively. Two methods of calculating net excitatory potential are then compared in Table 6. If a purely associative theory of drive is used (Figure 7), then net excitatory potential (E) should decrease with deprivation time, as shown in Table 6 when $E = H - I$. If, on the other hand, deprivation brings generalized energizing effects (D) as well as drive stimuli (S_D), and if $E = H \times D - I$, then more inhibition will have to be generated to stop the response occurring at 12 hours deprivation, and net excitatory potential should reach a maximim at some period of deprivation between 4 and 12 hours (Table 6).

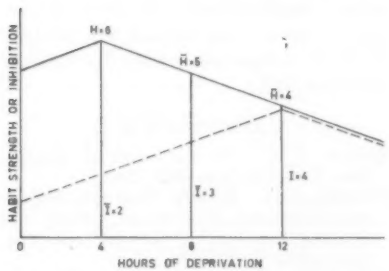

FIG. 7. The application of the Perkins-Logan theory to a hypothetical situation in which a given response is reinforced at 4 hours deprivation of food but not reinforced at 12 hours deprivation. (It is assumed in this instance that the effects of drive are purely associative.)

TABLE 6

ALTERNATIVE DERIVATIONS OF NET EXCITATORY POTENTIAL FROM VALUES ARBITRARILY ASSIGNED IN FIGURE 7

	Hours of deprivation		
	4	8	12
Habit strength (H or $\bar{H}$)	6	5	4
Drive strength (D)	2	4	6
$H \times D$	12	20	24
If $E = H - I$			
I equals	2	3	4
E equals	4	2	0
If $E = H \times D - I$			
I equals	12	18	24
E equals	0	2	0

CONCLUSIONS

Taken collectively, the results of the present experiments confirm the Perkins-Logan theory of stimulus-intensity effects in terms of differential conditioning and are incompatible both with Hull's original formulation and the modification proposed by the author. Thus the feature of the stimulus situation which seems to be directly responsible for the phenomena under study is the relative disposition

of the eliciting stimulus and the background or intertrial stimuli on some common dimension of generalization. The Perkins-Logan theory is also meaningful in the sense that it permits the generation of a number of predictions which remain to be tested. The denial of Hull's stimulus-intensity dynamism postulate and the verification of the differential conditioning theory, however, does not jeopardize the general principles of the type of behavior theory which he and his co-workers established. On the contrary, the differential conditioning theory follows directly from some general features of Hull's behavior system, based as it is on the accepted principles of excitation following reinforcement, inhibition following nonreinforcement, stimulus generalization, and the algebraic summation of excitation and inhibition. What Perkins and Logan have done is imply that an *assumption* of stimulus-intensity dynamism is redundant within the general framework of Hull's behavior theory, and the empirical data are in keeping with this implication. The third and fourth experiments described in this paper went one step further, however, in showing Hull's assumption of a multiplicative relationship between stimulus intensity and drive strength to be invalid as well as redundant.

SUMMARY

Following an outline of the Hullian and Perkins-Logan theories of stimulus-intensity effects, and a distinction between the role of these effects in learning and performance, four experiments were described which tested the relative merits of the two theories in situations limited to effects on performance. In the first experiment it was found that, with stimulus offset as a CS, response strength late in training was directly related to the strength of the offset stimulus. Because this finding could be readily accounted for by a modification of Hull's stimulus-intensity dynamism postulate couched in terms of degree of stimulus change as the effective variable, regardless of the direction of the change, the second experiment was designed to detect any negative transfer occurring with a shift from onset to offset or vice versa in the one group of subjects. Negative transfer did occur and this outcome was taken to be incompatible with the modification of Hull's postulate and as further evidence for the Perkins-Logan theory in terms of differential conditioning. In the third and fourth experiments tests were made of a deduction from the latter theory to the effect that there should be an additive relationship between stimulus strength and drive strength in their joint determination of response strength, contrary to Hull's assumption of a multiplicative relationship. The two studies provided evidence for an additive relationship. It was concluded from the series of experiments as a whole, that the Perkins-Logan theory is to be preferred to Hull's treatment of stimulus-intensity effects. Some other implications of the former theory were also considered, mainly through comparisons with discrimination learning.

REFERENCES

BROWN, J. S. *The motivation of behavior.* New York: McGraw-Hill, 1961.

CASTANEDA, A. Reaction time and response amplitude as a function of anxiety and stimulus intensity. *J. abnorm. soc. Psychol.*, 1956, **53**, 225–228.

CHAMPION, R. A. Motivational effects in approach-avoidance conflict. *Psychol. Rev.*, 1961, **68**, 354–358.

ESTES, W. K. Stimulus-response theory of drive. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1958.* Lincoln: Univer. Nebraska Press, 1958. Pp. 35–68.

GRANIT, R. *Sensory mechanisms of the retina.* New York: Oxford Univer. Press, 1947.

GRANT, D. A., & SCHNEIDER, D. E. Intensity of the conditioned stimulus and strength of conditioning: I. The conditioned eyelid response to light. *J. exp. Psychol.*, 1948, **38**, 690–696.

GRANT, D. A., & SCHNEIDER, D. E. Intensity of the conditioned stimulus and strength of conditioning: II. The conditioned galvanic skin response to an auditory stimulus. *J. exp. Psychol.*, 1949, **39**, 35–40.

GRICE, G. R., & SALTZ, E. The generalization of an instrumental response to stimuli varying in the size dimension. *J. exp. Psychol.*, 1950, **40**, 702–708.

HANSCHE, W. J., & GRANT, D. A. Onset versus termination of a stimulus as the CS in eyelid conditioning. *J. exp. Psychol.*, 1960, **59**, 19–26.

HULL, C. L. Stimulus intensity dynamism (*V*) and stimulus generalization. *Psychol. Rev.*, 1949, **56**, 67–77.

HULL, C. L. *Essentials of behavior.* New Haven: Yale Univer. Press, 1951.

HULL, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

LOGAN, F. A. A note on stimulus intensity dynamism (*V*). *Psychol. Rev.*, 1954, **61**, 77–80.

MILLER, N. E. Liberalization of basic S-R concepts: Extensions to conflict behavior motivation, and social learning. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959. Pp. 196–292.

NYGAARD, J. E. Cue and contextual stimulus intensity in discrimination learning. *J. exp. Psychol.*, 1958, **55**, 195–199.

PERKINS, C. C., JR. The relation between conditioned stimulus intensity and response strength. *J. exp. Psychol.*, 1953, **46**, 225–231.

RABEN, M. W. The white rat's discrimination of differences in intensity of illumination measured by a running response. *J. comp. physiol Psychol.*, 1949, **42**, 254–275.

SCHWARTZ, M. Conditioned-stimulus variables in avoidance learning. *J. exp. Psychol.*, 1958, **55**, 347–351.

SPENCE, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychol Rev.*, 1937, **44**, 430–444.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

SPENCE, K. W. *Behavior theory and learning.* Englewood Cliffs: Prentice-Hall, 1960.

SPENCE, K. W., HAGGARD, D. F., & ROSS, L. E. UCS intensity and the associative (habit) strength of the eyelid CR. *J. exp. Psychol.*, 1958, **55**, 404–411.

WALKER, E. G. Eyelid conditioning as a function of intensity of conditioned and unconditioned stimuli. *J. exp. Psychol.*, 1960, **59**, 303–311.

(Received March 3, 1961)

Psychological Review
1962, Vol. 69, No. 5, 450–466

CONDITIONING THROUGH VICARIOUS INSTIGATION

SEYMOUR M. BERGER[1]
Indiana University

Personality theorists and social psychologists generally recognize that the emotional responses of one person (performer) may elicit emotional responses from another (observer). When these emotional responses are similar, the relationship between the performer and observer is described as *empathetic,* or one of *identification.* If the relationship is one of *envy,* the performer's pleasure elicits an unpleasant emotional reaction from the observer. In a *sadistic* (malicious joy) relationship, the performer's displeasure elicits a joyful response from the observer. In all of these relationships, the observer's emotional response is dependent upon the emotional response of the performer. As social phenomena, empathy, envy, and sadism require that the observer experience the emotion of the performer *vicariously,* in the sense of perceiving it, as a prelude to the observer's own emotional response.

The purpose of this paper is to provide a conceptual framework within which to study empathy, envy, and sadism as determinants of interpersonal behavior. For example, empathy may account for certain forms of altruism, envy for the loss of friendship, and sadism for malicious mischief. Before proceeding to these problems, however, it is necessary to identify a basic mechanism in the arousal of the observer's emotional response.

Vicarious Instigation

If an observer responds emotionally to a performer's unconditioned emotional response (UER) then vicarious instigation has occurred.[2] The performer's UER requires more precise definition. It is meant to refer to the performer's emotional state, following the presentation of an unconditioned stimulus (UCS), *as perceived by the observer.* The actual state of the performer is not directly observable. The observer must infer the performer's UER from observable events in the environment. Typically, the observer will rely upon the UCS and the performer's overt unconditioned response (UCR) as cues to the performer's UER. Consequently, the performer's UCS and UCR are important determining factors in vicarious instigation.

The performer's UER need not actually occur, the observer may infer it (incorrectly) from the situational context. It is only necessary that the observer perceive that the performer's UER has occurred, and respond emotionally, for vicarious instigation. A mother, for example, may react with fear and anguish when she sees her child fall, only to find that the child is unhurt and undisturbed. In this example, the mother's emotional response (ER) was dependent upon her perception of what had happened to her child (the UCS) rather than the child's reaction (UCR). She misperceived the child's emotional state in this case, yet she was vicariously instigated; she inferred the child's UER from the

[1] The author wishes to express his indebtedness to at least two colleagues, Marvin Levine and Leon H. Levy, for their encouragement, many helpful suggestions, and critical reading of the manuscript.

[2] Throughout this paper, the term *unconditioned* refers to dependable S-R relationships which serve as a basis for conditioning. This usage is in agreement with the definition employed by many writers in the area of conditioning (e.g., Hilgard & Marquis, 1961, pp. 46–47). In this procedural sense, the term unconditioned may refer to either learned or unlearned S-R relationships.

UCS only. Had she been in another room and heard the child cry (UCR), the mother might have inferred that her child was in pain (UER) without knowledge of the precipitating event (UCS), and yet would have been vicariously instigated. While the performer's UCS and UCR are important cues for vicarious instigation, both are not always necessary for vicarious instigation.

Pseudovicarious Instigation

The simple contiguity between the performer's UCS and UCR and the observer's ER is not sufficient evidence for vicarious instigation, however. Since the observer's ER follows the performer's UCS and UCR, the observer may appear to be responding to the performer's UER when in actuality he is responding to some other stimulus component in the situation. As indicated above, vicarious instigation is dependent upon the performer's UCR or the UCS. In some situations, however, the observer may respond to the UCS only, so that the performer's UER is superfluous. An observer may respond with fear to the firing of a gun, regardless of whether the target is a bull's-eye or another human being. The observer's fear is not vicariously instigated in this instance—although it may appear to be if the target happens to be another human being.

Another possible source of pseudovicarious instigation is the performer's UCR. The performer's UCR may be a UCS which elicits an observer ER; in this case, the observer responds to the performer's UCR only, so that the performer's UCS and UER are superfluous. A sudden loud scream by a performer, for example, may elicit a fear response from the observer independently of his response to the performer's UCS or UER. Recent studies by Church (1959) and Murphy, Miller, and Mirsky (1955) have demonstrated how a rat's squeal and a monkey's presence, respectively, may serve as a conditioned fear stimulus.

The performer's UER may serve as a discriminative stimulus which indicates to the observer that he will receive the UCS that is presented to the performer. This is another source of pseudovicarious instigation. One child may fear that he will be spanked when he sees that his brother is being spanked for no apparent reason. It is not his brother's cries in response to the spanking that elicit the child's fear, rather it is the anticipation of being spanked himself.

The following study by Barnett and Benedetti (1960) illustrates a form of pseudovicarious instigation which may occur in conditioning. In this case, an extraneous stimulus in the situation served as a UCS:

After adapting the GSR of an observer to a buzzer, the authors had him watch a performer undergoing GSR conditioning, with shock as the UCS and the buzzer as the conditioned stimulus (CS). Even though most observers were not aware that the performer was being shocked, they responded with GSRs to the buzzer on test trials. Subsequent experimentation concerning the various stimulus components of the situation led the authors to conclude that what appeared to be "vicarious conditioning" (their term) involved responses to visible needle deflections of the galvanometer, even in the absence of the performer; the GSRs, in this case, reflecting changes in activation level following the occurrence of a stimulus change, i.e., needle movements.

All of these situations are not instances of vicarious instigation, although they would appear to be in casual observation. Obviously, therefore, anecdotal evidence for vicarious instigation is of dubious value, and an experimental analysis is required to eliminate the possibility of pseudovicarious instigation.

Concordant and Discordant Forms[3]

The definition of vicarious instigation requires that the observer respond emotionally to the performer's UER, but there is no restriction placed upon the nature of these emotional responses. As indicated previously, there may be a certain similarity or dissimilarity between the emotional responses of the performer and observer. Concordant vicarious instigation takes place when performer and observer responses are similar[4]; discordant vicarious instigation occurs when performer and observer responses are dissimilar. An examination of concordant and discordant forms of vicarious instigation provides a rudimentary classification system for empathy, envy, and sadism.

If discussion is limited to the elicitation of emotional responses which may be classified as either pleasant (+) or unpleasant (−), it is possible to specify the necessary conditions for empathy, envy, and sadism. The four possible combinations of pleasant and unpleasant emotional responses for performer and observer are presented in Table 1. It should be noted that throughout this discussion, the classification of the performer's UER is in terms of the observer's perception of it.

TABLE 1

COMBINATIONS OF EMOTIONAL RESPONSES FOR PERFORMER AND OBSERVER

Case	Performer's UER	Observer's ER
I	+	+
II	+	−
III	−	+
IV	−	−

Empathy may be defined by Cases I and IV; shocking the performer (−) may be painful (−) for the observer, for example. *Envy* may be defined by Case II; rewarding the performer (+) may instigate a feeling of depression in the observer (−). *Sadism* may be defined by Case III; shocking the performer (−) may elicit joy in the observer (+).

Although these combinations of performer and observer responses seem to identify the *necessary* conditions for characterizing empathy, envy, and sadism, the present scheme does not include the *sufficient* conditions for complete characterization.

Nevertheless, this initial effort suggests a useful direction to pursue in order to bring a wider variety of emotional responses within the scope of conditioning theory.

Conditioning

As a motivational concept, vicarious instigation has relevance to a variety of social behaviors acquired through classical and instrumental conditioning. In the discussion which follows, an attempt is made to identify two conditioning models in which social learning is a function of vicarious instigation.

[3] Although not specifically concerned with the problem of conditioning, Heider (1958, Ch. 11) has presented an analysis which describes one person's reactions "to the lot of the other." There are some parallels between the analysis of conditioning through vicarious instigation and Heider's presentation. The present analysis attempts to deal with similar problems, but in terms of conditioning models.

[4] Some authors (e.g., Asch, 1952, p. 171; Heider, 1958, pp. 279–282) distinguish between two forms of concordant emotional response. The observer may react concordantly with the performer's response to the unconditioned stimulus (UCS) and in a manner that is concordant with the way that the observer would react if he had been directly instigated; the observer may react concordantly with the performer's response to the UCS, but in a manner that is discordant with the way the observer would react if he had been directly instigated. In the former instance, the observer is reacting emphathetically, while in the latter instance the observer is reacting sympathetically. This distinction, while useful for some purposes, will be disregarded in this paper because it does not advance the present analysis.

CLASSICAL CONDITIONING PARADIGM: THE PERFORMER'S UER AS A UCS

In classical conditioning, in order for a response to become conditioned to a neutral stimulus (CS), the response must be elicited by the UCS. Any stimulus may serve as the UCS if it will elicit a given response with some degree of regularity. In vicarious instigation, the performer's UER serves as a UCS for the observer's ER.

The distinction between vicarious instigation and conditioning may be illustrated as follows: Haner and Whitney (1960) had an observer watch a performer who supposedly was being shocked each time a light was presented. The authors found that this situation elicited GSRs from the observers, the GSR size being positively related to the general anxiety level of the observer. But no test was made to determine whether the observer's GSRs became conditioned to the light alone. While Haner and Whitney call this phenomenon "empathic conditioning," the term vicarious instigation seems more appropriate.

In the classical conditioning model, vicarious instigation must occur in contiguity with a CS. Conditioning occurs when the CS comes to elicit the observer's ER in the absence of the performer's UER. While this model may be applied to a variety of social learning situations, an example of concordant vicarious instigation in the formation of attitudes may be found in this excerpt from Gordon W. Allport (1935):

> Even before he has an adequate background of appropriate experience, a child may form many intense and lasting attitudes toward races and professions, toward religion and marriage, toward foreigners and servants, and toward morality and sin. A parent's tone of voice in disapproving of ragamuffins who live along the railroad track is enough to produce an uncritical attitude in *the child who has no basis in his experience for the rational adoption of the parent's point of view* [italics added]. It frequently happens that *subsequent* experience is fitted into the attitude thus uncritically adopted, not . . . made the basis for the attitude. In such cases every contact is prejudged, contradictory evidence is not admitted, and the attitude which is *borrowed secondhand* [italics added] is triumphant (p. 811).

In terms of the model, the "ragamuffins" are the CS, the parent's UER is inferred from the disapproving tone of voice, and the child's uncritical negative attitude is the observer's ER. Once the emotional component of the attitude is conditioned ("borrowed secondhand"), the cognitive elements are added by subsequent experience.

Although attitudes are formed in a variety of ways, the present model provides a conceptual and experimental basis for studying the transmission of ready-made attitudes through observational learning in a classical conditioning framework. One can only speculate concerning the variety of social experiences that may be transmitted in this way.

INSTRUMENTAL CONDITIONING PARADIGM: VICARIOUS INSTIGATION AS A REINFORCEMENT

Vicarious instigation may serve as either a positive or negative reinforcement for observer responses in instrumental conditioning situations. Each type of reinforcement will be considered separately.

As a positive reinforcement, vicarious instigation strengthens the observer responses it follows. It is assumed that a positive ER is instigated in the observer. Therefore, pleasant empathetic feelings or malicious joy would be the basis for the positive reinforcement. Altruism, for example, is assumed to be based upon concordant vicarious instigation where both performer and observer emotional responses are positive. To illustrate, I may become more charitable when my

gift is received with obvious pleasure by the recipient. If perceiving the pleasure of the recipient is vicariously instigating, it may reinforce my charitable behavior. Of course, sources of pseudovicarious instigation would have to be controlled, since charitable behavior may be reinforced by approval I receive from the recipient or from my family and friends.

"Tattletale" behavior is an example of how a sadistic relationship may affect conditioning. One child may learn that he can have his rival punished by disclosing the rival's misdeeds to his mother. Presumably, the enjoyment elicited by this rival's punishment reinforces the tattletale behavior. Again, experimental analysis is necessary to eliminate pseudoeffects; e.g., the child may be seeking the approval of the mother and a raise in his own status.

Vicarious instigation may function as a negative reinforcement if responses are strengthened when followed by the removal or reduction of vicarious instigation. If it is assumed that emotional states have motivating properties then, following Miller's (1948) paradigm, vicariously instigated emotions may function to initiate learning; furthermore, the observer will acquire responses which reduce the emotional drive stimulus.

The performer's UER may instigate the observer's ER (assumed to be negative). The observer's ER is the drive stimulus which instigates other responses. If one of these other responses is followed by the removal of the performer's UER, vicarious instigation is terminated, and so is the observer's ER. The response which was followed by the termination of vicarious instigation will be strengthened.

This model seems applicable to situations when the vicariously instigated emotion is negative. Consequently, both unpleasant empathetic feelings and feelings of envy can function as drive stimuli in this model. Consider, for example, the starving beggar who comes to the door for a handout. The housewife may feel grief at the sight of the emaciated beggar and upon hearing his plea. She may be able to reduce her grief by shutting the door to remove the pitiful figure from her view. In this case, the beggar's state (performer's UER) aroused the housewife's grief (observer's ER) which was reduced by shutting the door. Presumably, the response of shutting the door was effective in reducing the feeling of grief and was reinforced. There are, of course, other potential sources of reinforcement in this type of situation. For example, the housewife may feel threatened by the presence of a stranger at her door, and may shut the door to insure her own safety. Clearly, sources of pseudoeffects must be ruled out before it can be concluded that the behavior was a function of vicarious instigation.

Envy may also function as a drive stimulus in this model of conditioning. One man may be envious when his rival is promoted, and attempt to reduce his feelings of mortification by subsequently avoiding all encounters with his rival. Again, it should be noted that an analysis for pseudovicarious instigation must be made (e.g., the man may be responding to his loss of the promotion rather than his rival's good fortune) before concluding that vicarious instigation is the basis for the behavior.

A number of writers (e.g., Berger, 1961; Campbell, 1961; Hill, 1960; Sechrest, 1961) have used the term *vicarious reinforcement* to describe learning phenomena. The present model identifies vicarious instigation as a possible basis for vicarious reinforcement.

AN EXPERIMENTAL DEMONSTRATION OF VICARIOUS INSTIGATION IN CLASSICAL CONDITIONING[5]

The foregoing analysis has employed anecdotal evidence and hypothetical examples to illustrate and support the conceptual framework. In light of the discussion of pseudovicarious instigation, an experimental analysis is imperative, yet lacking. Since vicarious instigation is crucial to the framework, demonstrations seemed in order. As a starting point, the author selected the classical conditioning model as a basis for demonstrating vicarious instigation in learning.

Although developed independently, the general setting for the experiments to be reported is similar to the situations used by Haner and Whitney (1960) and Barnett and Benedetti (1960). In this situation, an observer watches a performer (actually a confederate of the experimenter, who supposedly is being shocked every time a light dims. The dimming of the light is preceded by a buzzer. The GSRs of the observer are recorded as a measure of his ER. Everytime the light dims and the performer is supposedly shocked, he jerks his arm from the inductorium.[6] The buzzer was the CS. The performer's UER was inferred from the dimming of the light and electric shock (performer's UCS) and the performer's arm movement (performer's UCR). Presumably, if the observer responded with a GSR on conditioning trials, vicarious instigation occurred. It was necessary, of course, to introduce pseudovicarious instigation controls.

The previous discussion of pseudovicarious instigation suggested three general sources of pseudovicarious instigation: The performer's response, per se; the performer's UCS, per se; and the observer's anticipation that the UCS will be presented to him.

EXPERIMENT I

In an initial experiment, an attempt was made to obtain information concerning the effectiveness of the experimental situation for vicarious instigation, and to determine if the instigation of observer GSRs by apparent shock to the performer was due to the performer's response, per se. The experiment contained two conditions:

1. Shock-Movement (S-M)—the observer was told that the performer would be shocked and the performer reacted with an arm movement when apparently being shocked.

2. No Shock-Movement (NS-M)—the observer was told that the performer would make a voluntary arm movement at a given signal (light dimming), and that the performer was not being shocked.

If no differences were found between these conditions, then it would appear that observers were responding solely to the performer's movement, and that the performer's UER was not a determining factor. If differences were found, where S-M observers were more responsive, then some support would be provided for the phenomenon of vicarious instigation.

Fifteen subjects were assigned at random to each condition. The performer (a male confederate of the experimenter who appeared to be just another subject recruited for the experiment) was the same for all

[5] The author wishes to express his appreciation to Indiana University for an equipment grant in support of this research, and to Neil Schneiderman, Susan Kroener, and Patricia Stewart who were the performers and scorers in these studies. The experimenter for Experiment I was the author, Patricia Stewart was the experimenter in Experiment II, Susan Kroener was the experimenter in Experiment III.

[6] The performer jerked his arm rapidly, as a twitch. The performer tried not to make other movements, or change facial expression.

subjects. In order to avoid any systematic bias in the performer's behavior, the instructions regarding the different conditions were read to the observer before the performer appeared for the experiment. For observers in the S-M condition, the experimenter explained that another subject would be shocked in the last part of the experiment, and that he would be a few minutes late because the experimenter's assistant was testing the other subject to see what level of shock he could take; for observers in the NS-M condition, the performer's initial absence was explained on the basis that the experimenter's assistant was giving the other subject practice in making the arm movement. Instructions, which were common to both conditions and did not reveal the specific nature of the condition, were read aloud when performer and observer were present together.

The experiment was described to the observer as an attempt to determine whether different kinds of stimulation (buzzer and light diming) produced different types of skin resistance changes. S-M observers were told that the performer would be shocked every time the light dimmed in order to determine the effect of shock on the performer's GSR. The observer was reassured that he was a control subject and would not be shocked at any time. NS-M observers were told that the performer would try to jerk his hand off the apparatus in front of him every time the light dimmed to see if the movement affected his GSR. The observer was told that he was a control subject and that no one would be shocked.

Although the experimenter appeared to be taking readings from both the performer and observer, only the observer's GSRs were recorded. After an initial adaptation period, in which the buzzer was presented for 1.0 second followed immediately by dimming of the light for 1.0 second, the performer's hand was strapped to the inductorium in front of him. It was only during this latter phase of the experiment that the performer made arm movements.

There were 13 trials during this phase of the experiment. Ten trials were used for conditioning, with 3 test trials interspersed in the same position for all subjects. The intertrial interval was 15 seconds.

The method of moving averages with blocks of three trials was used for all curves presented in this paper, using the number of observers giving a GSR on given trials as the dependent response measure. For purposes of statistical analysis, the total number of GSRs for each observer, for a given set of trials, was used as the dependent response measure in all experiments.

Three criteria were used to identify a GSR response:

1. The movement of the recording pen had to be in the direction of lowered resistance.

2. The movement had to be at least as large as twice the thickness of the recording pen line.

3. The movement had to occur within 4 seconds following the termination of the buzzer.

The records were scored independently by the experimenter and the confederate, although both had some knowledge as to which conditions given observers had been assigned. The scorers agreed on 94% of the trials concerning whether or not a GSR had occurred; disagreements were resolved by re-examination and discussion by the scorers.

The criterion for adaptation during the first phase of the experiment was two successive no response trials; all observers were given a minimum of four adaptation trials, however. The adaptation curves over the four trials are presented in Figure 1. While it is apparent from Figure 1 that subjects in the S-M condition were less responsive than subjects in the NS-M condition, a comparison of the total number of individual responses over the four adaptation trials for each condition by the Mann-Whitney U test revealed no reliable differences ($U = 81.5$).[7]

The relative effectiveness of the two conditions to instigate observer GSRs may be examined in the "conditioning" section of Figure 1. The total scores for individual observers in the two conditions (on trials in which the confederate responded) were compared by the Mann-Whitney U test. S-M observers appear to be more responsive

[7] Two-tailed tests are used throughout this paper. The Mann-Whitney test was employed in Experiment I because the comparison groups differed significantly in variance (uncorrectable by usual transformation techniques) and, at the same time, appeared to have differently shaped distributions. The data for Experiments II and III were more amenable to analysis of variance according to the criteria suggested by Boneau (1960).

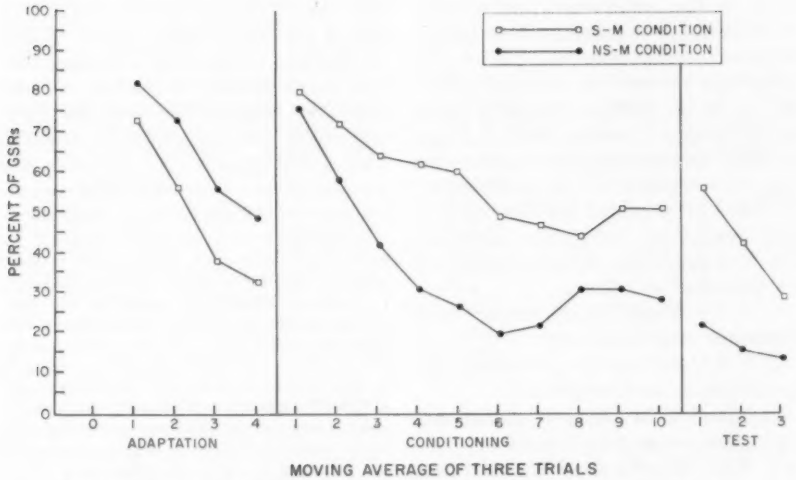

FIG. 1. Percentage of observers giving GSRs during adaptation, conditioning, and test trials (Experiment I).

than NS-M observers, but the difference is not reliable ($U = 70.5, p < .10$).

The total number of GSRs over the three test trials was obtained for each observer and the Mann-Whitney U test was used to compare the differences between the conditions. The "test trial" section of Figure 1 shows that S-M observers gave more responses on test trials than NS-M observers. The total score differences were reliable at less than the .05 level ($U = 61.5$).

A post-experiment questionnaire revealed that most observers gave descriptions of the performer's situation which were in agreement with the experimental instructions. Two NS-M observers reported that the performer might be receiving a shock. One observer in each condition reported some expectation of receiving shock.

These results lend support to the notion that conditioning can occur through vicarious instigation. While the arousal effects of the arm movement are apparent from the data, there is some tendency for GSR instigation to occur more frequently when the observer is told that the performer is being shocked. Furthermore, the frequency of conditioned responses (CRs) is reliably greater in the S-M condition. Conditioning was not solely a function of the performer's response, but was dependent upon the instructions concerning shock.

EXPERIMENT II

The encouraging results of Experiment I testified to the utility of the experimental setting in studying vicarious instigation. Yet, Experiment I contained only two pseudovicarious instigation controls; i.e., the NS-M condition, for performer response, and the instructions to the observer that he would not be shocked. A third control is necessary to determine whether the performer's UCS per se, will account for the differences obtained. It is conceivable that S-M observer's were affected solely by the instructions concerning shock, and that the performer's response was not a contributing factor; or, that the apparent presence of elec-

tricity in the inductorium had arousal effects which were independent of the performer *being* shocked. Since the differences between the movement conditions for instigations was not reliable in Experiment I, a complete 2×2 factorial design involving shock and movement was employed in this experiment. This design provided for the various pseudovicarious instigation controls. The four conditions of this experiment are identified as follows:

1. S-M—Shock to performer, and performer arm movement.
2. S-NM—Shock to performer, but no performer arm movement.
3. NS-M—No shock to performer, but performer arm movement.
4. NS-NM—No shock to performer, and no performer arm movement.

The S-M and NS-M conditions were the same conditions used in Experiment I. The S-NM condition is a "UCS control" for the instructions concerning shock and an electrified inductorium; the performer made no response. The NS-NM condition was another UCS control; the performer's hand was not placed on the inductorium during the conditioning phase of the experiment. She was told that she would not be shocked, and should not touch the inductorium when the light dimmed or else she would be shocked. This condition was used to assess the arousal effect of an electrified inductorium without shock to the performer.

Twenty-four subjects were assigned to each condition so that each condition contained an equal number of males and females.

A female confederate acted as the performer in this experiment. In addition to the different instructions for the various conditions, there were some procedural changes. In Experiment I, an attempt was made to use a "blind" performer by having the critical instructions for the different conditions read to the observer before the performer entered the laboratory. This procedure was abandoned in this experiment for two reasons: the performer in Experiment I frequently could identify the condition being used by the kinds of overt responses the observer made (e.g., smiling, looking away when the performer jerked, apparent boredom, etc.), and the performer in this experiment had to know what condition she was in so that she could behave appropriately.

The criteria for scoring GSRs were the same as in Experiment I, except that the pen

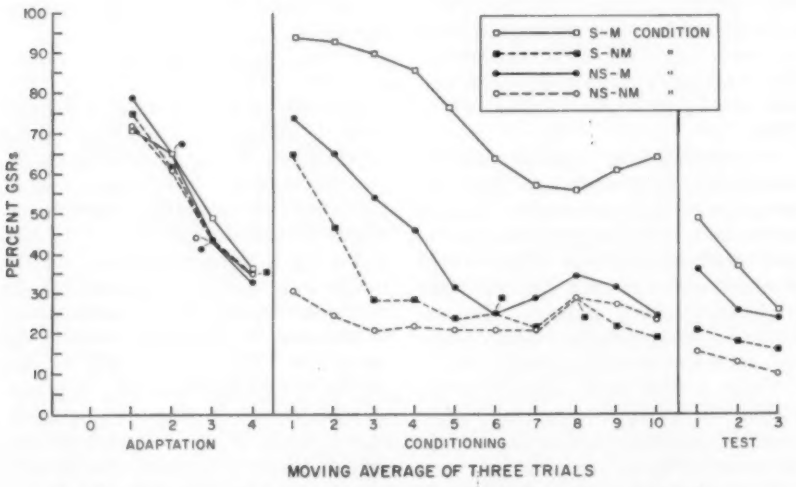

FIG. 2. Percentage of observers giving GSRs during adaptation, conditioning, and test trials (Experiment II).

movement in the direction of lowered resistance need only have been greater than the thickness of the pen line (rather than twice the thickness). This change represented an attempt to increase the sensitivity of the GSR measure, particularly with regard to CRs, which generally occurred relatively infrequently in Experiment I. Scoring was done independently, and without identification of the condition to which individual observers were assigned. The scorers agreed on 96% of the trials concerning the occurrence of GSRs; disagreements were resolved by reexamination and discussion by the scorers.

No differences in observer GSRs were observed between conditions during the adaptation period, as may be seen in Figure 2.

Table 2 summarizes the analysis of variance performed on the total scores for observers in the different conditions over the 10 conditioning trials.

Since the purpose of this study was to determine whether the differential effect of shock instructions found in Experiment I would occur without performer arm movement, the Shock × Movement interaction in the analysis of variance is particularly relevant. This interaction indicates that the differential effect of shock instructions depends upon whether or not performer arm movements are present. There is a significant difference between the frequency of GSRs given by S-M observers and the frequency of GSRs given by observers in the NS-M condition ($t = 4.33$, $df = 46$, $p < .01$); this difference is not significant between the observers in the S-NM and NS-NM conditions.

The main effect for Movement reveals that performer arm movements are effective elicitors of the GSR, regardless of shock instructions. The t between S-M and S-NM observers is 5.89, $df = 46$, $p < .01$; the t between the NS-M and NS-NM observers is 2.92, $df = 46$, $p < .01$.

The main effect for Shock is primarily due to the difference between the S-M and NS-M observers, the difference between the S-NM and NS-NM observers being small and not significant.

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE OF GSR INSTIGATION SCORES

Source	*df*	*MS*	*F*
Shock instructions	1	94.01	16.79**
Performer movement	1	219.01	39.11**
Sex of observer	1	4.60	—
Shock × Movement	1	38.76	6.92*
Shock × Sex	1	3.01	—
Movement × Sex	1	7.58	1.36
Shock × Movement × Sex	1	2.99	—
Residual	88	5.60	

* $p < .025$.
** $p < .001$.

The t tests reported above show that observers in the S-M condition differ from observers in the NS-M and S-NM conditions. The S-M observers also differ from NS-NM observers ($t = 7.35$, $df = 46$, $p < .01$).

The differences between conditions over conditioning trials may be seen in Figure 2.

The results of Experiment I suggested instigation was not solely a function of the performer's arm movement. This finding is strongly supported by the results of this experiment; a very reliable difference was found between the S-M and the NS-M conditions. In addition, this study reveals that the instructions concerning shock to the performer (S-NM condition), or that electricity is being passed to the inductorium (NS-NM condition) are not sufficient in themselves to account for the frequency of observer GSRs in the S-M condition. It is apparent from these data that frequency of instigation depends upon the instructions concerning shock and the occurrence of the performer's response.

Figure 2 also presents the test trial data for observers in four conditions. An analysis of variance of the test trials

scores revealed no reliable differences in conditioning. The order of conditions, based upon the magnitude of the means for CRs, is in line with the order of conditions based upon the means for instigations. The frequency of CRs tends to be greatest in the S-M condition, although the difference between S-M observers and observers in the other conditions was not great enough to yield statistically significant results.

Although reliable differences were found between the conditioning results from the S-M condition and the NS-M condition in Experiment I, the differences between responsiveness on test trials for these conditions in this study were not reliable, but in the same direction.

A content analysis of the post-experiment questionnaire data attempted to answer two questions: Did the observer describe the performer's actions in accordance with the instructions he received? Upon questioning, did the observer report that he expected a shock during the last part of the experiment?

In the S-M condition, one observer reported that he did not think the performer was being shocked, while this report was given by eight observers in the S-NM condition. Judging from their reports, a large number of observers in the S-NM condition believed that the performer was receiving a very mild shock or no shock at all; this may account for the relatively low frequency of instigations in this condition. Many S-NM observers did not perceive that the performer was being shocked or made uncomfortable. Five observers in the NS-NM condition reported that they did not think the performer would be shocked if she touched the inductorium. The majority of observers, however, described the conditions of the experiment in accordance with the instructions given, the one major exception being the S-NM observers.

Most of the observers who reported that they expected to receive a shock during the last part of the experiment, apparently expected the shock on the first trial or two, and when it did not happen, they no longer expected to be shocked.

When the mean GSRs from observers who expected shock are compared with the general means for the relevant conditions, it appears that all means are essentially equal to or less than the general mean for the particular condition—except for the instigation mean of S-M observers. Of the two observers in the S-M condition who expected shock, one gave GSRs on 8 out of 10 conditioning trials, which is equal to the median number of responses given by observers in the S-M condition; the other observer gave GSRs on all 10 trials, which is equal to the score of one-third of the observers in the S-M condition. Of the 24 observers in each condition, 2 S-M observers, 3 S-NM observers, 5 NS-M observers, and 5 NS-NM observers reported some expectation of receiving shock. Consequently, even though a small number of observers in each condition reported that they expected to be shocked, there is nothing in their data to suggest that this is a significant factor in their GSR responsiveness.

While there is general agreement between these findings and the results of Experiment I, the differences between the S-M and NS-M conditions with respect to instigation and conditioning were not consistently as large in both studies. When the probabilities of the observed differences are considered together, there is good support for the phenomenon of conditioning through vicarious instigation. Yet, it appears that some refinements in measuring

techniques might improve the reliabilities of the differences.

EXPERIMENT III

There is some possibility that the instructions and performer reactions resensitized the observer's response to the buzzer during the second part of the experiment, so that the measures of instigation and conditioning were confounded with a resensitized response to the buzzer, per se. Furthermore, the use of only three test trials may have been insensitive to some of the conditioning that occurred. For these reasons, the measurement technique was changed to conform to the recommendations of Stewart, Stern, Winokur, and Fredman (1961), in which the latency of the GSR was used to identify response measures. By taking into account the latency of the GSR to a stimulus, these authors found it possible to obtain separate measures of responses to the CS and UCS, and to identify conditioned anticipatory responses, for each trial. With this technique, therefore, it is possible to have each presentation of the stimuli serve as a test trial for conditioning, and simultaneously obtain measures of responses to the CS and to the UCS. The purpose of this experiment was to determine the utility of the latency technique for conditioning through vicarious instigation. Only the S-M and NS-M conditions were examined in this experiment, using the same instructional procedures as reported for Experiment II.

Some procedural changes with regard to presentation of stimuli were introduced to accommodate the latency technique. During adaptation, only the buzzer was presented for 6.5 seconds. The intertrial interval was varied at random (during adaptation and conditioning) from 20 to 60 seconds. During conditioning, the buzzer was presented for the same length of time as in adaptation. The dimming of the light for 1 second followed the termination of the buzzer by .5 second. As usual, the performer jerked her arm when the light dimmed.

For scoring purposes, the stimulus interval was divided into three 4-second units.[8] The first 4-second unit represented the response to the buzzer; the last 4-second unit, measured from the end of the light dimming period, represented the response to the performer's movement; the second 4-second unit represented a conditioned anticipatory response.

The criteria for identifying GSRs was the same as used in Experiment II, except that the pen movement in the direction of lowered resistance had to occur within one of the 4-second intervals, and the movement had to indicate an accelerating change rather than a decelerating one; this latter criterion seemed necessary to distinguish GSRs initiated in one scoring period from one initiated in the next period, on a given trial. Again, scoring of the records was done independently, without identification as to which condition the observer had been assigned. The scorers agreed on 93% of the cases, and resolved their disagreements by discussion.

The performer was a female in this experiment, but not the same person who was used in Experiment II.

Sixteen subjects were assigned to each condition so that each condition contained an equal number of males and females.

The adaptation for the two conditions was quite similar. Figure 3 shows the response frequencies to the buzzer, the second, and the third scoring units for both conditions over the first four adaptation trials. Since only the buzzer was presented during adaptation, responses occurring during the second and third scoring periods are assumed to be spontaneous.

When responses to the buzzer during conditioning are compared to the buzzer responses occurring during adaptation (see Figure 3), it is apparent that observers are resensitized in the conditioning period. An analysis of variance based upon total responses

[8] The 4-second unit was used in Experiments I and II for identifying GSRs to the stimuli. It is also in agreement with the general findings of Stewart, Stern, Winokur, and Fredman (1961, p. 63).

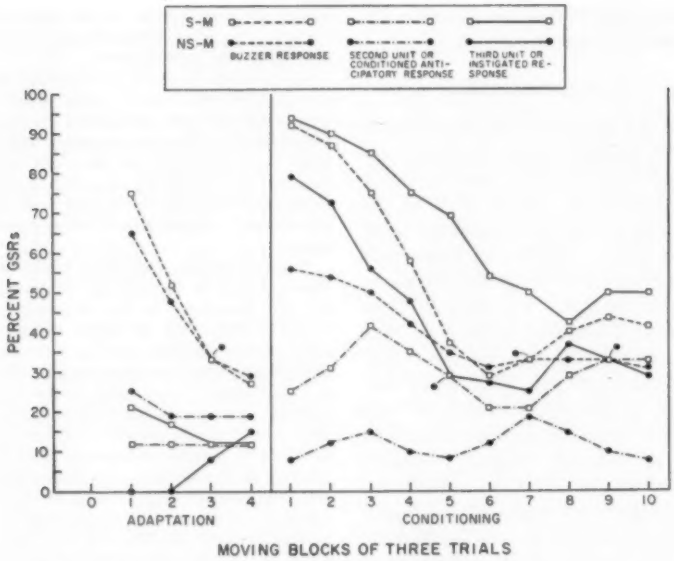

FIG. 3. Percentage of observers giving GSRs during adaptation and conditioning (Experiment III).

to the buzzer during conditioning revealed no reliable differences between the two conditions.

The curves presented in Figure 3 for instigated responses have the same general shape as the curves for the comparable conditions in the two previous studies; S-M observers are more responsive than NS-M observers. The analysis of variance based upon total instigations for each observer shows that this difference is reliable (see Table 3).

S-M observers gave more conditioned anticipatory responses than NS-M observers, as may be seen in Figure 3. The analysis of variance, summarized in Table 4, shows that the difference in conditioned anticipatory response totals for observers in the two conditions is reliable.

In this experiment, the experimenter asked each observer at the end of the experiment if he expected to receive a shock. Only one observer in each condition reported that he expected shock.

These findings provide additional support for the phenomenon of conditioning through vicarious instigation. The general level of observer instigation was similar to that found for comparable conditions in Experiment II, but a general increase in the number of conditioned GSRs was noted in Experiment III, where anticipatory re-

TABLE 3

SUMMARY OF ANALYSIS OF VARIANCE FOR INSTIGATIONS

Source	*df*	*MS*	*F*
Shock instructions	1	40.50	5.36*
Sex of observer	1	8.00	1.06
Shock × Sex	1	0.00	
Residual	28	7.77	

* $p < .05$.

TABLE 4

SUMMARY OF ANALYSIS OF VARIANCE FOR CONDITIONED ANTICIPATORY RESPONSES

Source	*df*	*MS*	*F*
Shock instructions	1	26.28	9.16*
Sex of observer	1	2.53	—
Shock × Sex	1	1.53	—
Residual	28	2.87	

* $p < .01$.

sponses were examined on all 10 trials by the latency criterion.

DISCUSSION

Considering all three studies together, there is a good deal of consistency in the findings with regard to the S-M and NS-M conditions. Although the differences between these conditions for instigations and conditioned responses was not consistently reliable in Experiments I and II, the trends in the differences were always in the same direction. With particular reference to the CR data, the finding in Experiment III that observers in both conditions are resensitized to the buzzer suggests that this factor may have tended to mitigate the reliability of the differences.

The reader may be concerned with the possibility that the differences between conditions may be influenced by different base levels of resistance, produced by the different treatments. The base levels of resistance (in ohms) were generally lower for S-M observers than for observers in the control conditions in these studies, although frequency of GSRs was greatest in the S-M condition. Since there is a positive relationship between GSR size and base level of resistance (Haggard, 1945), this would suggest that the effect of base level differences tended to decrease the differences in frequency of GSRs between the S-M and control conditions, according to the GSR criteria used. Consequently, if base level differences did influence the results of these studies, it was in the direction of minimizing vicarious instigation effects.

It should be noted that the present experimental situation *underplays* the instigation of the performer. This was done in order to keep the S-M and NS-M conditions plausible and yet comparable. Presumably, the level of instigation (and consequently, conditioning) would be higher if the performer's behavior was changed so as to provide greater evidence of pain. Actually, the present level of instigation and conditioning is high enough to provide evidence for the phenomenon, and yet low enough to provide a basis for testing the facilitating effects of a variety of variables.

In all of the data, observer GSRs tend to occur less frequently during the last half of the conditioning period than the first half. The explanation for this change is not entirely clear. One possible explanation is that GSR adaptation occurs because the observer becomes habituated to the specific situation as a result of repeated presentations of the stimuli, as Sears (1933) has shown. In addition, relatively short intertrial intervals tend to produce rapid GSR adaptation to repeated stimuli (Coombs, 1938). It also is possible that observers pay less attention to the stimuli as the experiment progresses. For example, some observers in the S-M condition reported that they could not bear to watch the performer being shocked and looked away after the first few trials; observers in the NS-M condition may have become bored with, and less attentive to, the repetitious movement.

It might also be argued that vicarious instigation phenomena have a transitory quality, the effect diminishing with successive exposures to the performer's situation as the observer be-

gins to exercise some internal controls over his ERs. An adult audience witnessing a movie, for example, may be emotionally aroused as they watch their favorite star fall off a cliff for the first time; but their level of arousal will probably be lowered after witnessing the same scene over and over again. Individuals can exercise control over their emotions when they can anticipate the occurrence of an emotionally arousing situation; although, considering the number of times a child can sit through the same movie with apparently undiminished enthusiasm, it would seem that children are less capable than adults in exercising such control.

Since it cannot be determined, from the data reported, whether the adaptation effects noted in these studies are a function of the characteristics of the GSR or a function of the characteristics of vicarious instigation phenomena generally, it would seem that a different measure of vicarious instigation is needed in order to avoid this confounding.

At the same time, some measure is needed to identify the nature of the observer's ER. While the reader may be tempted to think of these experiments as reflecting some sort of observer *empathy*, there is no basis for this interpretation. Some observers may be empathizing and some observers may be enjoying the performer's apparent pain. Research is presently underway which employs the evaluative scales of the Semantic Differential for this purpose. It may be possible to measure conditioning (as Staats & Staats, 1957, have done), with the performer's UER as the UCS for the observer. If this research is successful, it not only should provide a basis for determining how attitudes and values may be conditioned through vicarious instigation, but it should be possible to examine, independently of the GSR, whether adaptation of vicariously instigated responses occurs.

The general experimental situation used in these studies appears to be a useful one for measuring conditioning through vicarious instigation. It should be noted that the overall findings lend support to the concept of vicarious instigation as defined in the introduction to this paper: the instigation and conditioning of the observer's GSR cannot be attributed to the performer's movement, per se, or to the electrified inductorium, per se, or to the instructions concerning shock, per se. It is the apparent shock of the performer (and by inference, the performer's UER) which is the effective instigating stimulus for observer response.

Vicarious Instigation and Conditioned Emotional Response

Vicarious instigation is certainly not a new problem in social psychology. The phenomenon has been identified many times by other terminology. McDougall (1908) referred to the phenomenon as the "sympathetic induction of emotions." Floyd H. Allport (1924) reacted strongly against this instinct doctrine and sought to bring the phenomenon within the scope of conditioning theory.

> A theory far more plausible than that of instinctive induction of emotion . . . is the principle of *conditioned emotional response.* It may be illustrated by the panic which seizes all the persons in a throng when a few of them show signs of terror. Granting that the true cause for alarm has been seen by only the original few, we have here a case of fear aroused by the process of sympathy. The explanation, according to the present theory, is as follows: we have been previously terrified in company with others and so have our fear emotion transferred to characteristic attendant stimuli, such as cries and visible expressions of the emotion in those about us. We now react at once to

the sight of fear in others by a fear response of our own. Here the conception of sympathetic induction loses its force. We fear not merely because we see the expression of fear in others; but because we have learned to read these expressions as signs that *there really is something to be afraid of*. It is not fear induced from others that we experience, but *our own* fear of dangerous situations which has been conditioned by social stimuli (p. 235).

Allport's (1924) analysis argues against vicarious instigation, as defined in this paper; i.e., the expressed emotions of others are simply discriminative stimuli for subsequent presentation of the UCS to the observer. His analysis may hold for *some* situations, but it does not seem applicable to the experiments just reported. Observers were reassured that they would not be shocked, and therefore had nothing to fear for themselves; and most observers reported that they were not concerned for themselves. Furthermore, the results of the experiments show that the arousal effects are greatest for S-M observers, when the performer's situation is made to appear most unpleasant; but the number of S-M observers who apparently expected to be shocked is equal to or less than the number of observers who expected shock in the less arousing conditions.

While these arguments are not conclusive, they suggest that vicarious instigation is either a special case of conditioned emotional responding or that vicarious instigation depends upon a different learning principle. The resolution of this issue is not imminent, since further clarification of the nature of conditioned emotional responses and vicarious instigation is required before suitable arguments can be formulated.

Vicarious Instigation in Relation to Experimental Operations

The use of verbal instructions to manipulate the independent variables, and a post-experiment questionnaire to check the subject's acceptance of the instructions, limits the application of the designs used for experimental analysis of vicarious instigation to certain classes of human subjects. It is clear that these experimental operations are not appropriate for work with lower animals, young children, or with mental retardates who have little verbal facility.

One possible solution to this problem is to substitute specific training histories for the instructions. Observers first may be given direct experience with the condition that the performer will undergo, and then be exposed to a performer in that condition. Subsequently, the observer may be tested to determine if his behavior was affected by exposure to the performer. Although it seems feasible to provide different training histories for each experimental and control condition used in studies of vicarious instigation, experimental evidence is needed to demonstrate the suitability of the training technique for investigation of vicarious instigation phenomena.

To some extent, the present definition of vicarious instigation has a certain conceptual inadequacy. In order to identify conditioning through vicarious instigation, it is necessary to show that sources of pseudovicarious instigation do not account for the behavior in question. From an experimental viewpoint, the presence of a particular factor is more easily demonstrated than its absence. Until a more positive statement can be formulated concerning the necessary and sufficient conditions for vicarious instigation, it may be anticipated that the present definitions will tax the experimenter's ingenuity for setting up experimental situations which eliminate sources of the pseudoeffects mentioned previously. In spite of this inadequacy, the con-

ceptual framework has experimental value.

CONCLUSION

This paper offers a conceptualization of conditioning through vicarious instigation and provides some experimental support for the phenomenon. The importance of the problem is made evident by the paucity of experimental evidence concerning the role of the phenomena of empathy, envy, and sadism in the acquisition of interpersonal behavior. The analysis of pseudo-vicarious instigation dramatizes the inadequacy of anecdotal evidence in this area.

No systematic attention has been given here to the personality and social determinants of vicarious instigation, or to the consequences of vicariously instigated conditioning for social interaction. Heider (1958), for example, identifies a wide range of common sense principles which could be incorporated into the conditioning models presented in this paper. One of the virtues of these models is that they permit the perceptually oriented theorist to apply his principles in a conditioning framework and allow the S-R learning theorist to test his principles with a wider range of stimuli.

REFERENCES

ALLPORT, F. H. *Social psychology.* New York: Houghton Mifflin, 1924.

ALLPORT, G. W. Attitudes. In C. Murchison (Ed.), *A handbook of social psychology.* Worcester, Mass.: Clark Univer. Press, 1935. Pp. 798–844.

ASCH, S. B. *Social psychology.* New York: Prentice-Hall, 1952.

BARNETT, P. E., & BENEDETTI, D. T. A study in "vicarious conditioning." Paper read at Rocky Mountain Psychological Association, Glenwood Spring, May 1960.

BERGER, S. M. Incidental learning through vicarious reinforcement. *Psychol Rep.*, 1961, **9**, 477–491.

BONEAU, C. A. The effects of violations of assumptions underlying the *t* test. *Psychol. Bull.*, 1960, **57**, 49–64.

CAMPBELL, D. T. Conformity in psychology's theories of acquired behavioral dispositions. In I. A. Berg & B. M. Bass (Eds.), *Conformity and deviation.* New York: Harper, 1961. Pp. 101–142.

CHURCH, R. M. Emotional reactions of rats to the pain of others. *J. comp. physiol. Psychol.*, 1959, **52**, 132–134.

COOMBS, C. H. Adaptation of the galvanic response to auditory stimuli. *J. exp. Psychol.*, 1938, **22**, 244–268.

HAGGARD, E. A. Experimental studies in affective processes: II. On the quantification and evaluation of "measured" changes in skin resistance. *J. exp. Psychol.*, 1945, **35**, 46–56.

HANER, C. F., & WHITNEY, E. R. Empathic conditioning and its relation to anxiety level. *Amer. Psychologist*, 1960, **15**, 493. (Abstract)

HEIDER, F. *The psychology of interpersonal relations.* New York: Wiley, 1958.

HILGARD, E. R., & MARQUIS, D. G. *Conditioning and learning.* (2nd ed. rev. by G. A. Kimble) New York: Appleton-Century-Crofts, 1961.

HILL, W. F. Learning theory and the acquisition of values. *Psychol. Rev.*, 1960, **67**, 317–331.

MCDOUGALL, W. *An introduction to social psychology.* London: Methuen, 1908.

MILLER, N. E. Studies of fear as an acquirable drive: I. Fear as motivation and fear reduction as reinforcement in the learning of new responses. *J. exp. Psychol.*, 1948, **38**, 89–101.

MURPHY, J. V., MILLER, R. E., & MIRSKY, I. A. Interanimal conditioning in the monkey. *J. comp. physiol. Psychol.*, 1955, **48**, 211–214.

SEARS, R. Psychogalvanic responses in arithmetical work. *Arch. psychol., N. Y.*, 1933, No. 155.

SECHREST, L. Vicarious reinforcement of responses. *Amer. Psychologist*, 1961, **16**, 356. (Abstract)

STAATS, C. K., & STAATS, A. W. Meaning established by classical conditioning. *J. exp. Psychol.*, 1957, **54**, 74–80.

STEWART, M. A., STERN, J. A., WINOKUR, G., & FREDMAN, S. An analysis of GSR conditioning. *Psychol. Rev.*, 1961, **68**, 60–67.

(Received March 13, 1961)

Psychological Review
1962, Vol. 69, No. 5, 467–476

CONDITIONAL–OUTCOME CHOICE BEHAVIOR IN RATS[1]

FRANK A. LOGAN

Yale University

A conditional-outcome choice has the property that, for at least one of the alternatives, the outcome itself depends upon the subject's behavior subsequent to the choice. Because this type of situation is somewhat novel, it may be useful to begin with an example.

Figure 1 shows the floorplan of a maze in which a rat can run from the start (S) around one of two perimeter paths to the goal (G). If he turns to the right after leaving the start box, his path takes two additional right turns and he then arrives at a second choice point, shaped somewhat like a Y maze and exposing two swinging doors. One of these doors is white and the other black, and reward at the goal is given if the rat selects (say) the white door regardless of its position. That is to say, the second choice point contains a conventional visual stimulus discrimination problem. Alternatively, if the rat turns to the left after leaving the start, his path takes two additional left turns and, at the location corresponding to where in the other wing he encounters a discrimination problem, here one of the two guillotine doors is closed forcing him, half the time to the left, half the time to the right, through a grey swinging door into the goal. Typically, he is fed in this case with the frequency that he is earning reward by his discrimination performance in the other wing.

The important thing to recognize is that, for the first alternative described, the outcome cannot be specified in advance. Whether or not the subject gets reward depends upon what he does, in this case, upon whether he goes through the correct door. It is this feature that the subject has control over the outcome that characterizes conditional-outcome choices.

This type of situation has considerable interest in its own right. As a single but familiar example, a person selecting a profession cannot just consider various jobs with associated probabilities of payoffs because the payoff depends upon how he performs in his chosen field. How much one earns as a scientist, a doctor, or even a ditch-digger depends in part upon how good he is at that job. The outcome, in short, is conditional upon performance.

But beyond this intrinsic interest, conditional-outcome choices have an important relationship to the problem of response definition. The discussion of this relationship in the next section will also show the logic of the researches to be reported in subsequent sections.

Conditional-Outcome Choices and the Problem of Response Definition

Fundamentally, the problem of response definition is one of deciding upon a rule, or set of rules, stating the

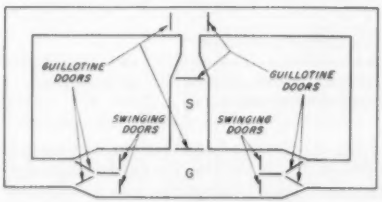

FIG. 1. Floorplan of the block-eight maze.

[1] Supported in part by the National Science Foundation.

conditions under which behaviors that differ in some way may nevertheless be aggregated into a single response class. All familiar response terms are aggregates of one kind or another. For example, there are a great many ways that a rat can get from one end of an alley to the other but these are all lumped together indiscriminately in the expression, the rat "ran down the alley."

The conventional solution of this problem seems to be that responses can be aggregated so long as they are not differentially reinforced (see, e.g., Guthrie, 1940; Hull, 1943). Thus, no distinctions are made among the various ways of running down an alley because the reward that is given at the end of the alley typically does not depend upon how the rat got there. This approach is schematized for a bar-pressing situation in Figure 2.

In the upper left box are identified four responses that can occur in a two-bar Skinner box: the rat can press Bar A with the left or right paw, or can press Bar B with the left or right paw. Of course, each of these is already an aggregate of different postures, different forces, and so forth, but they will suffice for the purposes of discussion. Now the typical procedure is to reinforce the rat differentially with respect to which bar he presses but not to reinforce differentially which paw he uses. Such a condition is indicated by the wavy arrows showing that the rat gets 8 pellets for each of the Bar A responses and 2 pellets for each of the Bar B responses. According to the rule that responses that are not differentially reinforced can be aggregated, the description can be simplified to that shown in the upper right box. In this case, information about which paw the rat used is completely suppressed and one simply refers to Bar A responses and Bar B responses.

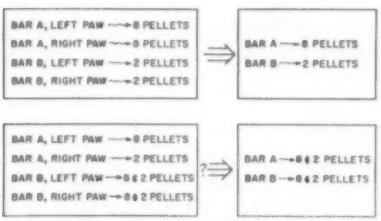
BAR A, LEFT PAW ⇝ 8 PELLETS
BAR A, RIGHT PAW ⇝ 8 PELLETS
BAR B, LEFT PAW ⇝ 2 PELLETS
BAR B, RIGHT PAW ⇝ 2 PELLETS
BAR A ⇝ 8 PELLETS
BAR B ⇝ 2 PELLETS
BAR A, LEFT PAW ⇝ 8 PELLETS
BAR A, RIGHT PAW ⇝ 2 PELLETS
BAR B, LEFT PAW ⇝ 8 ∮ 2 PELLETS
BAR B, RIGHT PAW ⇝ 8 ∮ 2 PELLETS
?
BAR A ⇝ 8 ∮ 2 PELLETS
BAR B ⇝ 8 ∮ 2 PELLETS

FIG. 2. Schematization of the responses in a two-bar Skinner box for an unconditional-outcome choice between the bars (top boxes) and a conditional-outcome choice between the bars (lower boxes).

The justification for this procedure is the same one that applies to the formation of any meaningful scientific concept, namely, that responses defined in this way enter into reasonably simple and consistent lawful relationships with other concepts (see, e.g., Spence, 1948). The substantial literature on choice behavior provides ample evidence of the usefulness of this procedure.

Suppose, however, that the reinforcement conditions were arranged as shown in the lower left box. In this case, the rat is differentially reinforced with respect to which paw he uses on Bar A, getting 8 pellets for left-paw responses and 2 pellets for right-paw responses. On Bar B, however, the reward is sometimes 8 and sometimes 2 pellets but this variation is not conditional upon which paw the rat uses.

Specifically, the question being raised is this: can the two Bar A responses be aggregated even though they are differentially reinforced? Can we suppress information about which paw the rat uses, and even ignore the fact that the reward is conditional upon which paw he uses? In short, can we, for purposes of predicting choice, refer to Bar A responses as shown in the lower right box? If so, then *that* response receives varied reinforcement, sometimes 8 and sometimes 2 pellets.

Again, the question of whether this step can be taken means, will a response defined in this way act like familiar responses that only aggregate nondifferentially reinforced behaviors? Will such a response obey the same laws and principles as determined under conventional conditions? One way of asking this question is to arrange the reward on Bar B so that the variation between 8 and 2 pellets there exactly matches that being received on Bar A as a result of trial-to-trial oscillation in the rat's behavior. If the aggregation is legitimate, then performance under these conditions can be predicted from familiar principles because the reward is not differential with respect to which bar the rat presses and he should, therefore, be indifferent between them when given a choice.

However, if the aggregation is not appropriate, if the familiar principles do not apply, or if new principles are required by conditional-outcome choices, then a preference might appear between alternatives that are otherwise equally reinforced. Specifically, for example, there might be something positive to a rat about having the reward under his control, in which case he would prefer an alternative such as Bar A where what he gets depends upon what he does. Or, with equal a priori possibility, such a condition could be aversive, in which case the rat would prefer not to have to make a differentiation on the way to the reward and so would choose an alternative such as Bar B. The following researches have been aimed at evaluating these possibilities.

METHOD

Subjects. Naive male rats were obtained from a variety of commercial sources at 90–120 days of age. They were housed in individual cages and were fed 12 grams of Purina laboratory chow daily 5–10 minutes after being run. Reward consisted of four 45-milligram Noyes pellets.

Apparatus. The studies of choice behavior when the outcome was conditional upon running speed were run in the double-alley maze that has been described in detail elsewhere (Logan, 1960). Briefly, it consists of a grey start box leading to two parallel alleys, one of which is white and the other black, both 4-feet long and equipped so that food reward can be delivered automatically according to a wide range of possible conditions.

The studies of choice behavior when the outcome was conditional upon a visual discrimination or a position response were run in the block-eight maze shown previously as Figure 1. This maze is grey throughout and is illuminated by six 7-watt bulbs spaced symmetrically around the maze and 18 inches above it. Food reward is given manually when appropriate by pulling a rod outside the maze that uncovers the food cup.

Procedure. Trials were run in blocks of four or six, one trial in each block being a free choice trial and the remaining trials being forced so as to insure that the rat went equally often in each direction. The sequence of forcing was arranged so as to avoid learnable sequences and to counterbalance alternation tendencies. The number of such blocks of trials run in a day varied from one to six in the various studies, and the intertrial interval was in some cases only the few seconds necessary to move the rat from the goal to the start box and in other cases was the time required to run the other concurrent rats. On every occasion that the same problem has been studied using procedures differing in any of these ways, the results have been closely comparable and there is no reason to believe that the results are peculiar to a particular aspect of the procedure.

One detail of the procedure that may be especially important concerns the technique for matching frequency of reward in the two alternatives. Half of the trials in each block were to each alternative and the rewards given in the unconditional alternative in any block of trials duplicated the rewards that the subject had earned in the conditional alternative on the preceding block of trials. Hence, there was a slight unavoidable lag in matching the rewards.

Outcome Conditional upon Running Speed

One way that the reward can be made conditional upon the rat's be-

havior is by the use of correlated reinforcement, as when reward is given differentially with respect to how fast the rat runs down an alley. A condition that we have studied fairly extensively (Logan, 1960) rewards the rat only if he takes at least 2.5 seconds between the first and last photobeams in the alley. The rat must leave the start box but then, rather than run down the alley in the 1–1.5 seconds it usually takes under uncorrelated reward conditions, he must dawdle for an additional second on the way. We have found that rats are continuously sensitive to the details of such conditions and that, after substantial training, they can attain as high as 75–80% reinforcement. In doing so, they typically develop some kind of ritualistic behavior, such as stopping and sniffing at the floor, or running with their heads over their shoulders which effectively slows them down. The question now is how to predict choice of an alley in which this type of condition prevails.

Correlated reinforcement versus no reward. Our first attack upon this problem used as an alternative alley one in which the rat was never rewarded. That is to say, the rat ran half of his trials in an alley in which he was rewarded only if he took over 2.5 seconds, and ran the other half of his trials in an alley in which he was not rewarded regardless of his speed. It was reasoned that, if there is something aversive about this type of conditional outcome, the rat might choose an alley in which he does not have to run slowly even though he does not get rewarded there. Of course, familiar principles imply a preference for the correlated alley since the occasional reward received there is better than nothing.

Three rats have been run on this problem and all three developed 100% preference for the correlated alley. This is all that was done to one of the rats. To the second rat, the conditions were reversed so that the correlated condition was placed in the black alley rather than the white one and nonreward was now in the originally correlated alley. This rat's performance quickly reversed indicating that the preference was strong and stable. The third rat was used to check on the following reasoning: if there is something aversive about a condition of correlated reinforcement, then the degree of aversion should be directly related to the length of time the rat is required to wait in order to get reward. Furthermore, since a longer cutoff also leads to a lower frequency of reward and hence reduces the positive aspect of the condition, any possible preference for the nonrewarded alternative should be more likely to appear the longer the cutoff. This reasoning is consistent with the finding of Herrnstein and Morse (1956) that pigeons would peck a key to terminate a signal indicating that a schedule requiring a long interresponse time was in force on another key. Accordingly, the cutoff for this rat was progressively reduced and his performance is shown in Figure 3.

The regular stepwise line represents the cutoff which was .4 (corresponding to 2.5 seconds) initially and which was decreased in stages to .025 (corresponding to 40 seconds) shortly before the rat died. The solid curve decreasing along with the cutoff is the rat's average speed in the correlated alley. Although speed decreased as the cutoff was lengthened, the problem was getting more difficult and the percentage reinforcement, which is shown in the dashed curve, dropped progressively. Choice, it can be seen, quickly attained 100% preference for the correlated alley and, with only a small decline near the end, remained there throughout the study.

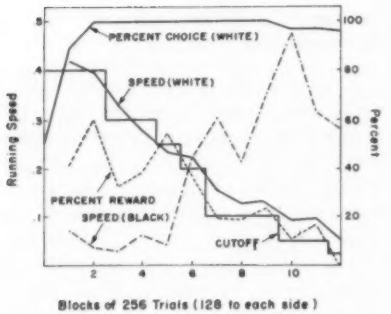

FIG. 3. Performance of an individual rat that was rewarded in the white alley only if his speed was below a cutoff that was progressively decreased during training. (The rat was never rewarded in the black alley.)

It might be noted in passing that speed in the nonrewarded alley, which is shown in the dot-dash curve, was somewhat erratic. Nevertheless, by the end of the study, the rat was consistently running appreciably slower in the correlated alley in an attempt to get reward than he was in the other alley for no reward. The fact that preference for the correlated alley persisted shows that, if there is anything aversive about having the reward conditional upon running speed, the effect is not very large.

Correlated reinforcement versus matched frequency of reward. Our next attack upon the problem of choice when the rat must run slowly in one alley used as an alternative alley one in which the rat was given the same frequency of reward as he was obtaining in the correlated alley. This arrangement is biased in that the rat can get the reward sooner in the uncorrelated alley because he does not have to run slowly there. Accordingly, since the response takes longer in the correlated alley, familiar principles imply that the rat should prefer the alternative alley. However, if there is something positive about conditions to correlated reinforcement, then the rats might forego reward for an extra second in order to control its occurrence.

Three rats have been run under these conditions and all learned to perform adaptively in the two alleys and consistently to select the uncorrelated alley. The results of one of the rats are shown in Figure 4 in order to illustrate that, at just about that stage of training where an appropriate difference in speed between the alleys began to appear, the choice curve, which had been falling presumably because what rewards were being received were obtained more quickly in the uncorrelated alley, showed a sharp but temporary rise. This same pattern was also clearly present in the curves of the other two rats and may reflect something positive about this type of conditional outcome. But the fact that all rats ultimately preferred the uncorrelated alley is consistent with familiar principles and shows that, if there is anything positive about having the re-

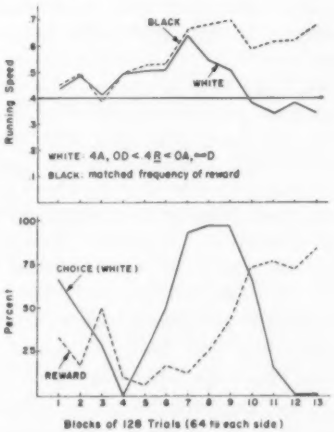

FIG. 4. Performance of an individual rat that was rewarded in the white alley only if he took over 2.5 seconds and was given a matched frequency of reward in the black alley.

ward conditional upon running speed, the effect is not very large.

Correlated reinforcement versus matched frequency and interval of reward. Our final attack upon this problem attempted to equate both the frequency and the interval of reinforcement in the two alleys. To do this, the problem in the correlated alley was changed somewhat, the rats in this case being required to run down to the entrance to the goal section of the alley and to wait there for at least 2 seconds before proceeding to the goal. If they failed to spend at least 2 seconds in that foot of the alley immediately preceding the goal, they were not rewarded. This restriction on the correct response was made so that the guillotine door in the alternative alley could be closed to detain the rats at the same distance from the goal and for a length of time matched to the time they detained themselves in the correlated alley. Since the frequency of reward was also matched, there is no apparent bias in the reward conditions favoring one alley over the other and there should therefore be no preference between them.

Four rats have been run under these conditions and, during the first 1,024 trials, all four developed and maintained 100% preference for the correlated alley, for the alley in which they timed their delay at the entrance to the goal section of the alley. However, when the conditions were reversed for three of these animals, only two reversed their preference within the next 1,024 trials, the third continuing to choose the same direction which was now the uncorrelated alley. Hence, there is some evidence for a positive aspect to an alternative in which the reward is conditional upon speed, but the effect is either not very large or not permanent.

Outcome Conditional upon a Discrimination Problem

A second way that the reward can be made conditional upon the rat's behavior is to use a visual discrimination problem. This situation has already been described, it being one in which, if the rat goes in one direction in the maze shown in Figure 1, he encounters a second choice point at which he must choose between a black and a white door, and if he goes in the other direction, he is forced through a grey door. Frequency of reward is matched in the two wings and hence familiar principles would lead one to expect the rat to be indifferent between them.

Black-white discrimination versus matched frequency of reward. Fourteen rats have been run with a black-white discrimination problem in one wing of the block-eight maze, and their combined performance is shown in Figure 5. There were wide individual differences in rate of learning the discrimination problem but the graph nevertheless correctly reflects the relevant features to be seen by examining all the individual curves. As the discrimina-

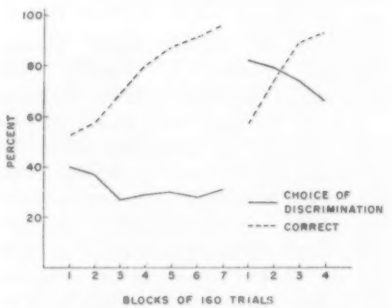

FIG. 5. Combined performance of 14 rats that encountered a black-white dsicrimination in one wing of the block-eight maze and were given a matched frequency of reward in the other wing. (The break in the curves indicates where the discrimination problem was placed in the opposite wing.)

tion problem was initially solved, a preference for the nondiscrimination wing appeared, only two rats choosing the discrimination wing on more than half the last 160 trials. But there was not a clear reversal in preference when the location of the discrimination problem was reversed. Hence, there is some evidence for a negative aspect to an alternative in which reward is conditional upon a visual discrimination, but the effect is either not very large or not permanent.

Dark-light discrimination versus matched frequency of reward. The direction of the effect with a discrimination problem was opposite to that obtained when reward was conditional upon running speed. There are several differences other than the nature of the task that might account for this discrepancy. Most obviously, there is a difference in difficulty because a rat on a correlated reinforcement condition in the alley initially gets reward very infrequently and never attains more than 75–80% correct, whereas a rat on a discrimination problem begins with 50% reward and ultimately attains essentially perfect performance.

In an attempt to explore this possibility, the discrimination problem in its reversed position was made more difficult for eight of the rats by decreasing the difference between the brightnesses of the two doors, generally by converging on an intermediate grey. This form of transposition was very good, the rats by-and-large continuing to choose the lighter or darker door depending upon the way they had originally been trained, but the difficulty level was in this way effectively increased as evidenced by a decrease to about 85% success. There was no observable effect upon choice as between the discrimination and nondiscrimination wings of the maze after 1,280 trials under these more difficult conditions, 59% of the choices in the last 160 trials being to the discrimination side.

Black-white discrimination versus dark-light discrimination. As a further test of the effect of difficulty of the task, a discrimination problem was then placed in both wings of the maze, one being the simple black-white problem and the other, a more difficult discrimination between two greys. Darker (or lighter) was correct on both sides. In order to avoid a preference developing between these alternatives because of a difference in reward based on a difference in percentage success, the frequency of reward in the simple discrimination was restricted to the same level that the rat was obtaining on the more difficult problem. There was no evidence of a preference after 1,280 trials under these conditions. The rats all continued to choose the direction they had been preferring during the previous phase of the experiment, which for some rats meant choosing the more difficult wing and for others, choosing the easier wing. Hence, problem difficulty, at least for experienced animals with matched frequency of reward, is not a critical variable in conditional-outcome alternatives.

Outcome Conditional upon a Position Response

A third type of situation in which the general problem of conditional-outcome choices can be tested involves the same maze as used for visual discrimination problems, but requires a position response at the second choice point. In this case, the swinging doors are all grey and in one wing of the maze, reward is conditional upon the rat going through (say) the left door. In the other wing of the maze, the turn is forced and reward frequency is matched to the percentage correct in the conditional alley.

One difficulty with this test of choice behavior when one alternative is conditional upon response selection is that the same type of response, namely, a position response, is also used to determine the rat's preference. If there is any anticipatory response tendency, it could contaminate the measure of preference; e.g., a preference for a left turn at the first choice point might appear because a left turn is correct at the second choice point rather than because that direction leads to (or away from) the position problem. A similar complication would have arisen in the previous cases had a stimulus discrimination or response speed differentiation been used as the indicator response. For example, a rat could encounter a black-white discrimination problem at a second choice point if he chooses white at the first choice point and not if he chooses black there, in which case the differential reinforcement of brightness at the second choice point might transfer to the first. It may be useful later to repeat these studies using different indicator responses.

However, there are several reasons for believing that this problem is probably not serious in the present instance. There is evidence (Logan, 1961) that learning is quite specific to its location in this apparatus, and unpublished research by students failed to detect an anticipatory response tendency. Furthermore, detailed examination of the present data did not reveal any consistent relationship between preference at the first choice point and the correct turn at the second.

Seven rats have been run on this type of problem and Figure 6 shows the results for four of them for whom location and correct position were counterbalanced. Only one rat showed a strong preference for the conditional-outcome alternative but the overall picture leans somewhat in that direction.

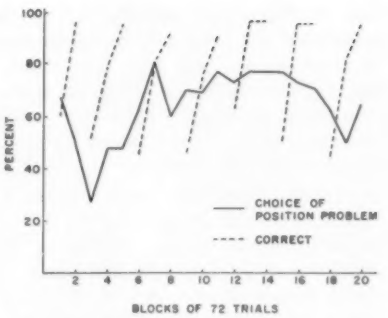

FIG. 6. Combined performance of four rats that encountered a position problem in one wing of the block-eight maze and were given a matched frequency of reward in the other wing. (Each break in the dashed curve indicates where the correct turn at the position problem was reversed.)

Hence, there may be a slight positive aspect to an alternative in which reward is conditional upon a position response.

Discrimination Problem versus Position Problem

Since the evidence suggested the possibility of something positive about a position problem and something negative about a discrimination problem, when each was compared with an unconditional outcome, it seemed reasonable that these effects might appear more clearly if they were pitted against each other. Toward this end, a black-white discrimination problem was placed in one wing of the block-eight maze and a position problem was placed in the other wing.

These conditions could not be used from the beginning of training because the two wings would be differentially reinforced. To avoid this difficulty, the rats were run through three phases preceding the testing phase. The first phase entailed running for 864 trials in just one wing to a discrimination problem or, for half the rats, a position

problem. The second phase entailed running for 864 trials in the opposite wing to the other problem. Because the position problem could be learned more quickly than the discrimination problem, the rats were given several reversal series in the position wing so that they received about the same number of reinforcements in each wing during the first two phases. The third phase entailed integrating the two problems, the rats being given an irregular series of 192 forced trials half in each wing with the appropriate problem in place. Four of the six rats made this transition to the integrated problem with only a few errors, but the other two had to be returned to a previous phase and then brought back to the integrated problem.

Hence, by the time of the fourth phase during which the customary sequence of free and forced trials was followed, the rats had solved both problems, gone each direction equally often, received comparable reinforcement frequencies in each wing, and currently were performing essentially perfectly on both problems. There is thus no basis in familiar principles to predict a preference.

None of the six rats showed a strong preference on the free trials during the fourth phase. Indeed, they were conspicuously indifferent, overall percentage choice of the position problem being 55%. Hence, there is no evidence that the nature of the problem makes a difference in conditional-outcome choices so long as frequency of reward is equated.

DISCUSSION

The reader may wish to emphasize those few instances in which we have obtained preferences that do not appear to follow from familiar principles. However, it now seems likely that the deviations encountered were either chance ones or the result of some artifact. Since the matching of rewards entailed a slight lag, any increase in frequency of reward appeared first in the conditional alternative and this difference could be sufficient to produce a weak preference for conditions of correlated reinforcement. The temporary preference away from a discrimination problem might simply reflect the fact that, during one stage of learning the rats had to pause to look back and forth at the discriminanda (see, e.g., Tolman, 1955), and the conditional alternative therefore took a moment longer to run. Certainly, if a powerful new principle were operating in conditional-outcome choices, its effect should have appeared more clearly in the various types of researches reported. Hence, the favored conclusion is that decisions in conditional-outcome choices can be understood on the basis of the same principles that govern nonconditional-outcome choices.

Of course, additional research is needed to establish the generality of this conclusion. Other procedures, particularly ones in which the conditional aspect is very prominent, may be more likely to detect a preference, and other organisms, particularly primates, may be more likely to reveal one. Furthermore, although a micromolar theory (Logan, 1960) implies that choice is the fundamental measure of differential response tendency, it is important to determine whether there are differences between the types of alternatives in measures such as resistance to extinction or in the effect upon them of variables such as drive. The results of these researches could deviate from the present findings but they would, at the same time, provide interesting comparative and methodological evidence.

Somewhat tentatively, then, it is concluded that there is no preference be-

tween a conditional-outcome alternative and one with the same frequency of reward uncorrelated with performance. This conclusion in turn supports the thesis that behaviors can be aggregated into a single response class even though there may be differential reinforcement among them.[2] The selection of a response depends upon the outcomes actually being encountered without regard to the fact that those outcomes may depend in some way upon how the subject performs subsequent to the choice. Accordingly, it is proposed that there are no restrictions within the conditions of reinforcement on the formation of response classes.

SUMMARY

A conditional-outcome choice is one in which the reward for at least one of the alternatives depends upon the organism's performance. A series of studies was reported in which rats were given a choice between two alleys, in one of which they had to run slowly, choose the correct brightness, or turn in the correct direction in order to obtain reward. It was concluded that rats are indifferent between such an alternative and one in which the same frequency of reward was given regardless of performance. This finding was interpreted as indicating that distinguishable behaviors can be aggregated into a single response class even though they are differentially reinforced.

[2] This thesis is not antithetical to the micromolar assumption that responses differing quantitatively are to be treated as different responses. The present position is that one can aggregate responses of different speeds into a single class for the purpose of predicting whether the rat will run down one alley or another, but the speed of running will not necessarily index the strength of that response. When different responses are aggregated into a single class, whether they are differentially reinforced or not, no statements can be made about differences among the responses within that class.

REFERENCES

GUTHRIE, E. R. Association and the law of effect. *Psychol. Rev.*, 1940, **47**, 127–148.

HERRNSTEIN, R. J., & MORSE, W. H. Escape behavior maintained by the removal of a stimulus correlated with an intermittent schedule of food reinforcement. *Amer. Psychologist*, 1956, **11**, 495.

HULL, C. L. *Principles of behavior.* New York, Appleton-Century, 1943.

LOGAN, F. A. *Incentive: How the conditions of reinforcement affect the performance of rats.* New Haven: Yale Univer. Press, 1960.

LOGAN, F. A. On the specificity of discrimination learning to the original context. *Science*, 1961, in press.

SPENCE, K. W. The postulates and methods of "Behaviorism." *Psychol. Rev.*, 1948, **55**, 67–78.

TOLMAN, E. C. Principles of performance. *Psychol. Rev.*, 1955, **62**, 315–326.

(Received March 20, 1961)

MICROFORM

All journals published by the American Psychological Association are being made available on MICROFILM or MICROCARD.

Psychological Review
American Psychologist
Psychological Bulletin
Psychological Abstracts
Contemporary Psychology
Psychological Monographs
Journal of Applied Psychology
Journal of Consulting Psychology
Journal of Educational Psychology
Journal of Experimental Psychology
Journal of Abnormal and Social Psychology
Journal of Comparative and Physiological Psychology

Available only in volume units; no single issues.

For MICROFILM, order from:
UNIVERSITY MICROFILMS, INC.
313 North First Street
Ann Arbor
Michigan

For MICROCARD, order from:
J. S. CANNER & COMPANY, INC.
Microcard Division
618 Parker Street
Roxbury 20, Massachusetts

CONTEMPORARY PSYCHOLOGY

. . . APA's journal of book reviews

- Critical reviews of the most recent books in psychology and related areas
- The Editor's page: comments on news in the publishing world, on criticism, on literary style
- A letters-to-the-editor section
- Instructional Media: reviews and comments on films, TV programs, workbooks, teaching machines

AMERICAN PSYCHOLOGICAL ASSOCIATION
1333 Sixteenth St., N.W.
Washington 6, D. C.